

DOCUMENT RESUME

ED 095 176

SP 008 356

AUTHOR Kauffman, Dan; And Others
TITLE Assessment of Teacher Preparation Programs.
PUB DATE Jun 74
NOTE 144p.

EDRS PRICE MF-\$0.75 HC-\$6.60 PLUS POSTAGE
DESCRIPTORS *Evaluation; Evaluation Techniques; Followup Studies; Historical Reviews; Literature Reviews; Models; *Teacher Education; *Teacher Programs

ABSTRACT

The introduction to this paper defines educational evaluation, offers historical details on this facet of higher education, and presents the subject in perspective against the broad spectrum of higher education. Section 2 presents a general overview of practices and literature on teacher preparation and educational evaluation, and section 3 reviews current practices. Section 4, entitled "A Model for Assessment of Teacher Preparation Programs," recommends that each institution set up a "teacher preparation assessment group" that would be responsible for the preparation and implementation of the assessments, the analysis of the data, and the upgrading of instruments, procedures, and analysis. In section 4, two assessment procedures are considered: interaction analysis and survey analysis. There follows a description of a sequence of the major activities to be completed in the survey analysis of teacher preparation programs. The paper concludes with references, a list of dissertations related to assessment procedure, and a directory of institutions associated with teacher preparation assessment programs. (JA)

ED 095176

ASSESSMENT OF TEACHER PREPARATION PROGRAMS

by

Dan Kauffman

Margaret Gonsalves

Eugene Knight

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

SP 0718 336

June, 1974

TABLE OF CONTENTS

	<u>Page</u>
LIST OF FIGURES	iv
 <u>Section</u>	
I. INTRODUCTION	1
II. TEACHER PREPARATION AND EDUCATIONAL EVALUATION: AN OVERVIEW	10
III. CURRENT PRACTICE	29
IV. A MODEL FOR ASSESSMENT OF TEACHER PREPARATION PROGRAMS	46
Introduction	46
Assessment Procedures	48
Interaction Analysis	49
Survey Analysis	51
Major Activity Sequence	55
1-2: Formulation of Project and Objectives	58
1-15: Documentation	63
2-4: Selection of Survey Design	65
4-6: Item Construction	71
6-7: Instrument Format and Coding	78
2-3: Selection of Sample	85
3-5: Selection of Sampling Frame	89
5-9: Design of Sampling Procedures	92
9-12: Determining Sample Size	95
7-10: Develop Prototype Instrument	100
10-11: Pilot Test and Revision	103
11-12: Development of Final Instrument	105
4-8: Selecting Statistical Analysis and Processing of Data	108
8-12: Procedures for Processing Returns	114
12-13: Initial Mailing	117
13-14: Keeping Tract of Returns-- Charting and Follow-up	118
14-15: Methods for Analyzing Returns	122
15-16: Preparation of Final Report	125

	<u>Page</u>
REFERENCES	127
DISSERTATIONS RELATED TO ASSESSMENT PROCEDURE	132
DIRECTORY OF INSTITUTIONS ASSOCIATED WITH TEACHER PREPARATION ASSESSMENT PROGRAMS	139

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. Major Activity Sequence Survey Analysis Assessment of Teacher Preparation Programs	57
2. Relationship of Sample Size to Error	95

SECTION I

INTRODUCTION

The unprecedented growth of higher education in the United States, the complexities born of overexpansion, and the censure to which this social institution has been subjected during the current decade have necessitated some searching self-study. Suggestions for a reexamination of roles and recommendations for innovation and change have been offered at all levels: institutions of Higher Education, despite some characteristic deliberation, have not been slow on the uptake of the new ideas. Leaders in the field, roused out of their complacency, have come to perceive some justification for the outcry against "traditional" and "antiquated" procedures. As a result, efforts are being made to review and to evaluate, to reform and to update. Indeed an era of university growth and rationalization in the history of the United States has given way to another--of definition, recognition, and evaluation.

The introduction to this paper purports to define educational evaluation, to offer relevant historical details on this important facet of higher education, and to present the subject in its proper perspective against the broad spectrum of higher education. Subsequent sections will thence focus on this significant aspect of the total educational process and its potential as a viable technique for curriculum improvement.

Alfred Schwartz and Stuart C. Tiedeman (1957) explain:

Evaluation is the process of making judgments and coming to decisions about the value of an experience. The

process consists of two elements: (1), a goal or objective for the experience to be evaluated must be set, and (2), some measure of amount, status or progress must be made. An evaluation of the experience then involves a carefully considered judgment as to the adequacy or effectiveness of the experience as measured in the light of the objectives set for it.

In education, evaluation is the process of judging the effectiveness or worth of an educational experience as measured against instructional objectives. Evaluation makes use of measurement, but is not limited to it, nor synonymous with it. Measurement never gives more than an answer to the question, "How much?" Evaluation, on the other hand, seeks an answer to the question, "Of what value is the measure of the amount, status, or progress when compared with the instructional objectives?"

.....
Measurement, by whatever means it may be accomplished, be it a carefully constructed standardized test in mathematics or a rating scale designed to measure home or personal adjustment, is a basic part of the evaluation process. But measurement is not enough. Measurement must be seen in terms of human values and goals. Evaluation, focused upon philosophically and psychologically sound objectives, and based upon the best measurements that can be secured, is a key to securing effectiveness in the total educative process [pp. 1-2].

Briefly, then, evaluation is more than a mere compilation of factual data. It "implies a system of values and decisions about values involving human judgment [Woodring, 1957, p. 62]." Evaluation requires decisions by human beings on accepted practices and their possible improvement.

The concept of evaluation is not new: its roots, in fact, lie deep in antiquity. The Spartans of ancient Greece used tests to measure physical prowess and the Athenians carefully selected candidates from the property class for training in their Ephebia in the martial arts and tactics. Will Durant (1954) relates that "China established--first tentatively under the Han, then definitely under the T'ang dynasties--its

system of examination for public office [p. 800]." And the Bible records that the Gileadites used the test of the ability to pronounce the Hebrew "shibboleth" to screen friend from foe.

At the higher education level, evaluation was in evidence in the earliest European universities which made their spectacular rise in the thirteenth century. At the University of Paris, the Responsions admitted the freshman to the university; the Determination, the origin of our Bachelor's degree, marked his graduation; the Inception declared him fit to be a master; and in Paris, as in Bologna, the jus oblique docendi conferred on him the right to teach anywhere.

The beginnings of the modern concept of evaluation were visible in the nineteenth century. W. H. Cowley (1970), in the context of the current impact of this concept, records that the first self-study was executed in 1825, when Harvard's governing boards "accomplished the most effective reformation ever made of an American college [p. 152]."

The Harvard Corporation, the Board of Overseers, and the faculty all appointed committees which met for two years. They redeemed the functions of the President, they established instructional departments, they instituted a primitive elective system, they sectioned classes in the modern languages on the basis of ability, they revised the college statutes, and they abandoned money fines for student misdemeanors. And who made the decision? The governing boards, stimulated by Professor Ticknor's petitions and precipitated by student unrest [p. 152].

Harvard's reform was infectious. Amherst proposed revision of her curriculum to include technological subjects and teacher training, and Williams in Massachusetts added the exotic with the introduction of

foreign languages. Under the German influence, men like Henry Tappan (1805-1881) and George Ticknor (1791-1871) advocated the study of the experimental sciences and even engineering.

These developments stimulated the Yale Corporation into reactionary activity and at the instance of the Governing Board, the Yale faculty delivered the Yale Report of 1828, "which served as the definitive justification of the college system as it then operated [Hofstadter & Smith, 1970, p. 252]." The country, however, was outgrowing the doctrine of intellectual discipline and the institution of the Land-Grant Act of 1862 and the incorporation of agriculture and the mechanical arts in the curriculum combined to produce the historic comprehensive American university.

With the opening of Johns Hopkins and the elevation of Daniel Coit Gilman (1831-1908) to the presidency of this institution, research became formally recognized as a function of the university, even as teaching was from its inception, and as public service, too, is today. With the added dimension of research, "the research point of view has come to dominate undergraduate education and has created the most serious conflict in American higher education [Cowley, 1970, p. 156]." Kenneth E. Eble observes, "A disproportionate attention to research is most often cited as the specific cause for the neglect of teaching [Cowley, 1970, p. 156]." But research is here to stay, and research has demonstrated that our efforts in education are experiments, which once tried, must be evaluated and improved for the progress of human society.

From the continent of Europe came the first impulse for our modern instruments of educational testing, measurement, and evaluation. The movement, however, soon developed a distinctly American pattern and to the contributions of Sir Francis Galton (1822-1911), Alfred Binet (1857-1911), and Wilhelm Wundt (1832-1920) were added the studies of the World War I period and the active interest of several individuals in colleges and universities.

Even earlier, however, a study had been conducted, which made a major impact on the field of education. Five volumes, under the series title Adventure in American Education (Aikin, Vol. I, 1942; Giles, et al., Vol. II, 1942; Smith, Tyler, et al., Vol. III, 1942; Chamberlin, et al., Vol. IV, 1942; Thirty Schools Tell Their Story, Vol. V, 1943) tell the story of the Eight-Year Study and the thirty schools that participated. Schwartz and Tiedeman (1957) comment:

The Eight-Year Study of students and curricula in thirty secondary schools provided an opportunity for the development of a full-scale program of evaluation. Of great significance to the field of educational evaluation was the fact that in this study an attempt was made to put an evaluation rationale into operation. The teachers, administrators, and research workers engaging in the task of evaluating the effectiveness of educational programs had to develop their own instruments as they sought to evaluate study skills, critical thinking, appreciation, and interests. The Eight-Year Study marked a turning point, for it showed that testing specialists had too long been concerned with the knowledge aspects of education and had not placed enough emphasis on the so-called intangible outcomes of the educative process. The study was also significant in that it provided a laboratory in which many of our present-day specialists in evaluation were able to study at firsthand the problem of evaluation in education [p. 8].

The period of World War II was one of unlimited experimentation, which laid a broader base for the process of evaluation. The augmented interest in the subject in the years during and immediately following the war is reflected in a spate of literature, which made a profound impact on education, psychology, personnel management, and other allied areas in which the process of evaluation is of significance.

Most of the action in education, however, was still at the lower levels, for innovation and change have ever been slow to penetrate the portals of higher learning. The universities, in fact, were the last to accept the New Learning of the Renaissance, but when they did, changes of far-reaching significance swept through both education and religion. Of American higher education, Frederick Rudolph (1968) remarks, "Resistance to fundamental reform was ingrained in the American collegiate and university tradition, as over three hundred years of history demonstrated [p. 491]." This conservatism, nevertheless, has permitted modifications over the years as circumstances have warranted.

In the twentieth century, a gradual and sometimes imperceptible erosion of traditional values in education has been taking place. In 1916, John Dewey (1916, pp. 271-292) emphasized a need for incorporating utilitarian studies in the curriculum. Since then, much attention has been diverted to vocational and occupational education and to experiential learning. The value of internship training is being increasingly acknowledged in the public sector, at the political level, and in educational administration; and the professional preparation of academic administrators has been a topic of considerable deliberation

in academic circles in recent years. Changing trends and changing values have made their demands on higher education, demonstrating the need for the evaluation of present programs and practices and for possible adjustment and realignment.

During the current decade, much of the new thinking has been unwittingly precipitated by yet another wave of student unrest. Rudolph (1968) observes, "While governing boards and faculties went about their business as usual, students quietly and unknowingly reformed the American college and university [pp. 484-485]." The same historian records:

A student at the University of Texas . . . gave expression to a distressingly widespread definition of education: "If a professor can make an evaluation of how much education he thinks the students received, then the student is also in a position to make an evaluation of how much education he thinks the professor gave him [p. 495]."

Censure of the old academic order and the insistence on revised methods and procedures have thus given the concept of educational evaluation a new turn. Originally employed to discover the strengths and weaknesses of students, evaluation today has become an instrument also to assess the value of curricula and the effectiveness of teaching. The complexity of establishing norms, however, in academic instruction, organization, and administration demands much thought, research, and experiment. In the field of education, objectives are sometimes intangible and performance and productivity often defy measurement. These facts, nonetheless, do not rule out the necessity for and the use of evaluation. The variegated pattern of individual characteristics and situations makes discovery and interpretation more difficult, but

nevertheless more challenging. Acceptance of the challenge has given research and experiment a new emphasis underscoring the need for evaluation.

Simultaneously with student disaffection, a general discontent has been manifest with Higher Education as it is currently offered. Determination of the quality and effectiveness of school personnel preparation programs has therefore become a major concern not only of colleges of education, but also of national bodies like the American Council on Education, the American Association of University Professors, and the Association of American Colleges, as evidenced by projects sponsored and reports published. Such efforts to appraise faculty and to assess the quality of programs in Higher Education will be dealt with in some detail in a subsequent section, as will other recent studies that reflect the growing interest in educational evaluation at the level of higher learning.

The achievement of objectives and goals, however, calls for more than efforts in isolation. A need exists for joint institutional effort and for cooperation among teacher training institutions and between such institutions and school systems; for the establishment of a common cause, and the development of a mechanism for measuring productivity and evaluating education, that will serve common ends and promote a common interest in the progress of human society. The complex nature of contemporary life, the turmoil and conflict in higher education, and an employment market in disequilibrium underscore the desirability of improved methods and procedures in teacher education. The absence, however, of unified

criteria as a basis for such improvement calls for more information on methods of evaluating teacher preparation programs.

This publication, therefore, is a plea for commonality of purpose and mission in the training of school personnel, in the assessment of generally accepted practices, and in the conceptualization and establishment of improved practices. An effort will be made to present an overview of educational evaluation and its impact on teacher preparation; to explore current practice; and to offer significant data on the strengths and weaknesses of teacher education programs as they are currently designed, in the hope that this will permit the establishment of a continuous cycle of program improvement as well as assist laymen in critical educational roles to relate their efforts to those of the professionals.

While it is not the aim of the authors to evolve the ideal or a specific program of teacher education or to develop the perfect design for evaluation in education, it is the purpose of this paper to assist educators in formulating judgments about the construction and improvement of programs for the preparation of school personnel and to provide an initial base of ideas in the development of a strategy for educational evaluation.

Finally, appended to this work is a directory of institutions which have made significant progress in assessing graduates, which, it is hoped, will generate the communication and contact that will promote further experimentation and development.

SECTION II

TEACHER PREPARATION AND EDUCATIONAL EVALUATION:

AN OVERVIEW

Particular attention to the preparation of school personnel was advocated as early as the sixteenth century by Richard Mulcaster (c. 1530-1611), English school master, educator, and author. This pioneer in education anticipated our modern efforts in teacher preparation by his recommendations for the formal training of teachers parallel to that in the professional schools for doctors and lawyers. Mulcaster was the first to suggest a training college for teachers. Of seven special colleges he planned at the university, one was for teachers. In Positions (1581), one of his two major pronouncements on education and the virtues of good sense and moderation in teaching, Mulcaster makes strong recommendations for "this careful provision for such a seminary of masters," setting forth reasons why "this trade requireth a particular college [Quick, 1896, p. 101]." Since then, over the centuries, much progress has been made in the institution of normal schools, teachers' colleges, and colleges of education, and in shaping professional values through formal education.

Today, colleges of education in the United States offer a mosaic of quality programs for the initial preparation and the continuing education of teachers and administrators at all levels and of other professional personnel engaged in the process of education in schools, colleges,

universities, and public and private agencies. For administrative purposes and to ensure specialization possibilities, many institutions have organized their courses of instruction into departments which offer a wide range of teacher education curricula. Related to different levels and facets of education, some of these departments are:

- Elementary Education
- Secondary Education
- Educational Administration and Supervision
- Counselor Education
- Educational Psychology
- Special Education
- Educational Technology and Library Science
- Higher Education--Administration
- Curriculum and Instruction

Supplementing the work of these departments are various bureaus, centers, special laboratories, offices, and other services, which directly assist the college and community. In the comprehensive American university, the College of Education plays no small part and the formal training of teachers who will guide the destinies of future American citizens receives due emphasis.

Of American education, Woodring observed in 1957:

Never before in history has a major nation provided so much education for so many for so long. Unfortunately the emphasis on quantity has not been accompanied by a similar emphasis on quality; we provide more education but it is not at all clear that we provide better education than other nations or an education that is nearly as good as we can provide with our vast resources [p. 4].

"Teacher education," said the same author, "is closely interrelated with all other problems of education and it is in the field of teacher education that problems of educational philosophy are brought most sharply

into focus [p. 3]." In 1955 the Fund for the Advancement of Education, an independent philanthropic organization instituted by the Ford Foundation in 1951, and later the Foundation's educational division, expressed concern over the "staggering problem of finding enough good teachers" to serve the needs of rising enrollments (Teachers for Tomorrow, 1955, pp. 17-19). The drive for new sources of teacher supply, however, has since overshot the mark and an overcrowded labor market has given problems of quantity a new and unexpected turn. A surplus availability of teachers has shifted the emphasis to quality and new and better ways of educating teachers are being explored.

New ideas in education and the extension of postsecondary education as a democratic right have introduced new philosophies and new methods of educational leadership. Consequent overexpansion in the area of Higher Education, spiralling costs, and the need for augmented federal and state support have given rise to the question of accountability. Pressures have been exerted for reforming the existing curriculum in colleges and universities and for the immediate implementation of innovative and effective educational programs. Among the recommendations made for change, Ann Heiss (1970) distinguishes "new arrangements of knowledge, new curriculum patterns, new methods or organization for learning, new interrelationships between disciplines, and new ethnic studies programs [p. 242]," against which our traditional methods have been considered outmoded, inadequate, and unsuited to present and future needs.

The concepts of universal opportunity and equal access, moreover, have accentuated the insistence on utility in college programs to suit the needs and aspirations of the individual. A growing emphasis has consequently been apparent on experiential learning and the need for the application of theoretical knowledge to real-world situations. Educators see the value of firsthand experience in the formal training of men and women to function in a variety of capacities.

The change in educational thought in the current decade has been marked by recommendations for the innovative and the distinctive. Suggestions have been made for a reorganization of the process of education to give the student more responsibility for his own learning and an opportunity for direct involvement with his environment. Among other recommendations are a reversal of the traditional teacher-learner roles with a view to making the student an active seeker of knowledge; variation of the lecture method of teaching; definition of academic goals and the means whereby they will be achieved; articulation between course objectives and means and departmental objectives and means; and modification of the approach to teaching to suit individual differences.

Alternative methods that have been tried are

supervising independent study, tutorials, undergraduate or graduate seminars, small group dialogues and discussions, workshops, student-faculty colloquiums, intergroup conferences, retreats, films and film-making, field study, or community-centered interdisciplinary studies [Heiss, 1970, p. 246].

The new forces have made their impact on teacher education, even as they have in other areas. However, lack of data on the effectiveness

of the various teaching methods precludes any conclusions about the ideal educational situation. It is at this point that the process of evaluation attains significance. In the trial and error process in which we have been involved, amid the outcry for innovation and change, we would be well-advised not only to examine our dependence on the past and the traditional, but also measure and evaluate our success with the distinctive and the unique. The emergence of new forces may not necessarily mean the annihilation of the old. Indeed it is possible some of the traditional concepts in education will intensify against the thrust of the new. The need to gauge educational effectiveness in the face of such complexities establishes the rationale for the process of evaluation in education. Paul Dressel (1970) writes:

It should be evident that systematic evaluation of instruction is essential. There are at least three identifiable reasons why this is so. The first is that it is required for the recognition and reward of good instruction; for faculty members, like students, are motivated by tangible recognition as much as or more than by personal satisfaction. The second is to provide knowledge and understanding which will make it possible to improve instruction and the learning by students which is the reason for providing instruction. The third, which represents an interest on the part of psychological and educational researchers who are primarily interested in the nature of learning and the facilitation of it, involves research on instruction, altogether apart from whether this research will be used to improve the process or not. In fact, however, such research is of little consequence unless it does contribute to improvement. Thus all three reasons ultimately condense into one: the improvement of the educational process [p. 2-64].

Other educators vocalize, in equally convincing tones, their awareness of the need for educational evaluation. Paul Woodring (1957)

declares, "A continuing critical evaluation of our own institutions is an essential part of the democratic process [p. 13]." The democratic character of our colleges of education, then, makes it imperative for each organization "to assess its own strengths and weaknesses and thereby to improve its operations and educational programs [Mortimer, 1972, p. 6]." Robert W. Heath (1969) states, "Whether as a developmental function, as an aid to the practicing educator, or as fundamental research, evaluation of the interaction between a curriculum and its environment seems essential [p. 280]." And to the dictums of psychologists and researchers, Lewis B. Mayhew (1969) adds:

The establishment, operation, and evaluation of the curriculum ought to be one of the central responsibilities of college faculties and academic administration. It is the vehicle through which the institution seeks to make its most significant impact on the lives of students [p. 188].

Programs of school personnel preparation may be evaluated by making judgments about the program itself, by judging the competence of the teachers who graduate from the program, or by evaluating the learning of the students taught by these teachers. The third approach, though the most effective, is the most difficult. "It requires effective control of a vast number of variables," says Woodring (1957), "and it requires a great deal of time, for the influence of the teacher continues throughout the lifetime of the pupil, and many of the most important influences are not easily measurable [p. 62]." The difficulties inherent in this method have given the other two alternatives wider use.

Currently, however, we have progressed beyond measuring the achievement of students as a method of evaluation in education. The most common approach to the evaluation of instruction now is through student evaluation of their learning experience. Student evaluation is doubly effective in that it provides the teacher with appraisal of his teaching, textbooks, assignments, examinations, and other considerations, and the student himself with self-appraisal, for he, too, has some responsibility in making the learning process possible.

The approach is not without its flaws. A low student rating does not necessarily reflect poor classroom performance. But, as Paul Dressel (1970) says, it is "relevant to know what students think about a teacher and about a course, for when students appraise an experience as being poor they are unlikely to attain the optimal benefit from it [p. 2-65]."

Other current forms of program evaluation are peer and administrative evaluation and self-evaluation. In the matter of instruction, peer and administrative evaluation may be addressed to courses, their content and requirements, the instructional pattern and interdisciplinary alignment. Self-evaluation, on the other hand, prompted by student ratings and peer and administrative comment, is essential to improvement.

These methods of evaluation have their problems. Validity and reliability of ratings, for instance, can be clouded by personal bias and prejudice, but overconcern with such preconceptions would only defeat the purposes of evaluation.

In general, the goals of evaluation in education are improvement of the educational curriculum, clarification of the purposes of the academic program, and measurement of the achievement of these purposes. With the introduction of non-traditional programs and the emphasis on utility, other goals have been established, but, reiterating Henry S. Dyer (1972), the primary purpose of evaluation is basically "deciding on what to do next in order to make things go better in the future [p. 392]."

The concept of usefulness in education has gained ground to the detriment of a broad and general education. The humanities, which "once were the vital force that exercised spiritual and intellectual leadership, . . . today occupy a most equivocal position in the academic galaxy," says Gustave O. Arlt (1965, p. vi). The advocates of specialism contend that education must equip the student for specific employment to enable him to take his place in the mainstream of American society. Indeed the federal statutes which instituted the Land-Grant colleges and occasioned the rise of agricultural extension services and vocational education programs had implications that training in specific skills would enable learners to utilize these skills with optimal benefit to themselves and to society.

Specialism, however, fetters with ignorance, prejudice, and narrow provincialism. Discussing the problems of the educational philosophy of the 1920's and the 1930's and the trend toward teacher specialization, Woodring (1957) states:

The solution appears to lie in requiring a broad liberal education for all teachers and school administrations and at the same time greater attention to educational philosophy during the period of professional education so that each teacher may see beyond his own field.

The liberally educated teacher will see his subject in broader perspective. His aim will be not knowledge of a subject, but the wisdom which follows knowledge. He will teach his subject better because of his ability to see beyond it.

In no profession is liberal education more important than in teaching and in few does it play so large a part. The teacher's need for such a background is not dependent upon the subject to be taught or the age level of the pupils with whom he is to work [pp. 8-9].

The passage of time, moreover, an accelerating technology, and the current knowledge explosion continually relegate to obsolescence skills that once were specific. When jobs become complex, a need for general education manifests itself. It is the liberally educated man then who can think, sift, differentiate, deduce, choose, and make wise decisions independently. The process of liberal education begins early in life, continues throughout adulthood, and is never completed. Indeed specialized or vocational education, when undertaken, should rest on the broad base of a liberal education.

Changing trends, of necessity, make education a continuing process equipping the student with new skills through a lifetime and providing him the wherewithal to cope with new problems. But, as James Bryant Conant (1958) expresses it, "Neither the mere acquisition of information nor the development of special skills and talents can give the broad basis of understanding which is essential if our civilization

is to be preserved [p. viii]." Expertise in particular areas of study does not presuppose an adequate background for citizens of a free nation. A program which stresses the mere acquisition of knowledge and skills fails to touch 'man's emotional experience as an individual and his practical experience as a gregarious animal [Conant, 1958, p. viii]." It has no ties with the past, on which actually the present is built; it ignores the wisdom of the ages, from which we received our cultural heritage. History, art, literature, and philosophy are necessary for the preservation of all that was good and great and beautiful. The transmission of such learning, weighted and influenced by yesterday, links the past with the present and renders form and continuity to society.

Though liberal education must be given the highest priority in terms of educating the complete man, this emphasis must not ignore other necessary aspects. There is more than a modicum of truth in the demands that seek to project education into the future as much as to work it backwards. In view of the current technological advancement, it can hardly be denied that an introduction to computer science and electronics or other aspects of scientific and technological study are important in equipping the student to take up his responsibilities in a complex world.

Changing values in society and in life then make it incumbent upon institutions of teacher education to attune the preparation of school personnel to the new wants and interests. It would be expedient, however, to build future possibilities against that broad culture base

of the past on which the structure of the present stands. A sound program of teacher education will prepare the teacher for a lifetime career rather than for his first job or for a specific job. Teacher education should lay the groundwork that would support a variety of careers and on which education would continue long after graduation.

The process of evaluation in education is a measure of the achievement of objectives and of the possible need for redefinition or modification. It is a means of determining progress or stagnation in a fast changing world. Evaluation helps sift the outmoded and the obsolete from the relevant and the useful. It is a search for the right choice amongst a host of values and, as such, an invaluable means to decision-making which is so essential to teaching and administrative competency and to the effective operation of an institution of Higher Education in all its phases. Curriculum evaluation has also a diagnostic function in that the process helps identification of the regions of the overall program that need attention and possible rectification, though the actual remedy "must be sought in more curriculum development and experimentation at the project level [Forehand, 1971, p. 582]." Finally, in communication with a large and varied population of respondents, evaluation could be a vehicle of publicity for the institution which employs it. Paul Dressel (1971) observes:

in a more responsible sense, such evaluation may help in determining the relevance of the programs to social needs and, thereby, in documenting the responsiveness of the institution to society [p. 171].

When properly applied, educational evaluation thus holds a distinct place in the process of education. It becomes a matter of concern, however, when it assumes the form of a goal in itself. A further difficulty presents in the possible creation of a negative atmosphere of constant evaluation. Comparative evaluation programs, moreover, tend to promote uniformity of course content and methods of presentation.

Dressel declares:

There is no common set of criteria that can be used for all programs, and there is no single set of standards which can be applied to them all. If local standards are developed, they will have to be prepared for groups of programs with high degree of similarity rather than for all [p. 169].

In the so-called diversity of content and method of American education, educators perceived one of its great strengths. This "diversity," however, has now come under criticism and one wonders whether our system of education offers adequate meaningful choices to suit the backgrounds, abilities, and interests of the heterogeneous population interested in postsecondary education. Woodring (1957) declares:

The educational philosophy for a democratic nation cannot be created by one man or by a professional group. It must emerge from the convictions of the people and must stem from their mores, their folkways, their ethical beliefs, and their concept of the good life. In a diverse nation, it must allow for diversity; in a changing culture, it must provide for change. Yet it must have sound moorings. It cannot be merely a reflection of the current tides or the whims of an uncertain people; it must have deep roots in the past and possess the stability provided by those roots [p. 5].

Implementation of an institutional policy of evaluation may encounter an opposition of views and be balked by varying and intangible

objectives. Researchers must also face the reality that educators have little control over many of the external factors that affect individual learning. Indeed the need to define objectives poses the greatest problem in educational evaluation, for without a clear-cut purpose the process would be meaningless, even impossible. There is an apparent need to broaden our concepts on evaluation, to whet our instruments, to define performance criteria, to participate and to get actively involved.

It is essential, therefore, that faculties, administrators, and researchers perceive the process of educational evaluation as a necessary adjunct of the professional activities of those concerned with curriculum development in general and with programs pertinent to teacher education in particular; and that they give it the priority it calls for. As Woodring (1970) says, "In spite of the fact that projects in teacher education are, by their very nature, difficult to evaluate, the problem of evaluation must be accepted as a major responsibility of all experimental projects if we are to know the extent of their success [p. 62]." We need to overcome our fears that evaluation in education is intended as a threat and to develop a truly professional climate in which the continued course of curriculum development and improvement can be maintained. Success in the area demands clear definition of goals and objectives, the accomplishment of these goals and objectives, and suggestions for even greater achievement; competency in making wise judgments; and financial support. An equally important need exists for a concerted drive to explore common issues and problems and to promote

common ends. Progress beyond the merely experimental will necessitate a combined effort by universities and colleges of education, foundations, and sponsoring and professional organizations.

The evaluation of traditional educational programs has engaged the attention of educators and researchers for years. "Non-traditional programs," says Hartnett (1972), "have all of the problems of evaluation in the more typical settings, plus a whole set of difficulties all their own [p. 31]." Evaluation in non-traditional settings assumes another and different purpose, namely, the need to establish credibility. Hartnett elucidates:

In the first years of non-traditional programs especially, it will be essential to convince others who might be wary--other educators, employers, potential students, the public--that flexible, non-traditional programs can be rigorous and demanding, resulting in graduates of quality [p. 32].

The problems encountered in the evaluation of non-traditional programs are twofold, arising out of two distinct aspects of the process: the assessment of individual student growth and development and the evaluation of the quality and effectiveness of non-traditional curricula. The two are closely related and necessary for the establishment of credibility. "The problem is an immense one . . . ," says Hartnett, "for the notions of program efficiency and individual student quality, in certain respects, can be expected to be at odds with each other [p. 32]."

That standards in non-traditional programs can be inversely related to the number of entrants that actually graduate is borne out

by the results of the year-long Educational Testing Service evaluation of Open University programs at the University of Houston, the University of Maryland, and Rutgers, which suggest "that the program is not appropriate for the academically weak [The Arizona Republic, May 26, 1974, p. B11]." The study, recently concluded, was directed by Hartnett, in an effort to determine the effectiveness of the British concept of the Open University in non-traditional offerings by American colleges.

In the evaluation of non-traditional programs that include an instructional component, achievement measures could be based upon the objectives and features of the academic program offered at each institution. Here, however, the danger presents that such achievement measures might dictate the educational program of the students, rather than allow the educational experience to develop in line with their needs and aspirations. Non-traditional programs with no instructional component pose an even greater problem for student certification and consequently for the evaluation of program efficiency. In this event, the assessment of student development and the evaluation of program effectiveness may both have to rely on standardized achievement examinations to demonstrate that graduates of non-traditional programs are as proficient as students who graduate from more conventional programs.

The need is evident for greater flexibility in awarding recognition of merit and in structuring academic programs. Hartnett (1972) observes:

Essentially, this greater flexibility would be an expression of awareness that learning can and does take place in different ways for different people

and that to continue to provide only one form of instruction and recognition is to be wasteful of the country's richest resources.

This is not intended to imply that many institutions of higher education are not now performing excellent educational functions. "Traditional" is not meant to be pejorative, any more than "non-traditional" is necessarily positive. The point is that there should be both. Students should have options, at least far more than are now available [p. 37].

Developments over the recent past offer evidence of a growing belief that educational evaluation can make a definite contribution to the improvement of curricular design in general, and positive conclusions derived show particular applicability to the analysis and improvement of school personnel preparation programs. In August 1957, the Fund for the Advancement of Education, whose prime activity was the support of experimental programs which promised advancing education in American schools and colleges, published Paul Woodring's report New Directions in Teacher Education, "in the hope that this summary [of its activities during its first six-year period] will stimulate further reexamination about practices in teacher education, that it will lead to higher quality in teacher preparation, and finally that it will play a useful role in the advancement of education [p. viii]." In March 1958 the Commission on Instruction and Evaluation of the American Council on Education expressed an interest "in the relations of teachers and administrators because it is clear that, wherever the principles may lie, these are the forces which, together with the students, determine the quality and effectiveness of higher education programs [Ellis, 1958, p. vi]." Judging from classified advertisements in the press, planning and

research and evaluation are becoming a feature of program development in American colleges of education. Administrators are looking for technical skills and experience in the development and coordination of evaluation activities, in the development of evaluation designs, in directing the design and implementation of formative and summative program evaluation, and in the use of evaluation consultants. State Departments of Education are thinking in terms of comprehensive statewide planning and evaluation systems for elementary, secondary, and community college education. The long-range planning activities of the Academy for Educational Development, Inc., Washington, D.C., for colleges, universities, and state systems of higher education, include postdoctoral training in institutional research and evaluation. The intended audience are educators from all levels concerned with particular aspects of evaluation or in need of refresher courses as a result of rapid advancement in computer techniques. The sessions have been designed to develop skills in educational research, with particular emphasis on new practices and perspectives in educational evaluation and conceptualization.

Such efforts, however, are still sporadic and the functions of evaluation in education still remain to be developed and coordinated into an effective system. As Robert W. Heath (1969) records,

In many respects the systematic evaluation of curricula is only beginning to emerge as a recognizable field of educational research. Curriculum reform in recent years has grown out of attempts to (1) bring the modern conceptual and methodological status of subject-matter fields into the experience of students, (2) apply current pedagogical thinking to classroom instruction, and (3) use the educational process to

achieve social-ideological goals. Typically, curriculum evaluation has followed, rather than inspired, these changes [p. 280].

The exponents of educational evaluation have expressed disappointment with studies designed to evaluate curriculum. In 1966, for instance, D. A. Abramson (1966) recorded a "continuing paucity of studies which can serve as models for curriculum research [pp. 388-395]." In May 1971, Garlie A. Forehand (1971) attributed the lack of systematic progress in the area to the problem of comprehensiveness--comprehensiveness of goals, for example--and the absence of a clear definition of the meaning and purposes of evaluation. Heath (1969) aptly sums the reasons for the lack of enthusiasm for curriculum evaluation.

The instruments employed have frequently been insensitive to the most important effects of instruction. Conventional tests, rating scales, and questionnaires have often been more convenient than relevant. Studies of curricular effects have answered questions of incidental interest, while issues of central concern have been left to informal, intuitive judgment. Though educators and parents are aware of socioeconomic, motivational, attitudinal, and emotional differences among students, these antecedent variables have been generally ignored in curriculum evaluation. Too often curricula have been defined in terms of texts, labels, and catchphrases rather than detailed objective descriptions of the educational treatment. Also, resistance to rigorous evaluation of instructional programs has come from criticism-sensitive educators and from curriculum innovators who have heavy personal and professional investments in their products. Finally, the agencies that sponsor nationwide curriculum developments have failed to support impartial evaluation of the programs they promote [p. 280].

Far less research has been done in the area of teacher preparation programs. What has been attempted, whether of entire teacher

education programs or of segments of programs, has been scattered and inadequate. Most of the available studies have been conducted by individual institutions. The following section will outline some representative studies in curriculum evaluation in general and in the evaluation of school personnel preparation programs, both traditional and non-traditional.

SECTION III

CURRENT PRACTICE

As indicated earlier, some progress has been made in educational evaluation during the current century and a striving for excellence or at least an effort to meet minimum qualitative standards has been manifest in colleges and universities in this country. The need for the assessment of quality at both the graduate and undergraduate levels has been emphasized on a national scale as well as by individual institutions and organizations. The number of organizations formed to conduct curriculum research and evaluation reflects the interest that has been generated in this facet of the educative process. The Center of Instructional Research and Curriculum of the University of Illinois, for instance, assisted by the faculty and staff of the institution, has been responsible for a series of curriculum studies and several theoretical papers. The Center of Curriculum Studies at the University of Minnesota "represents a substantial commitment to curriculum research by a major university [Heath, 1969, p. 282]." Among other institutions that have a long and extensive history of interest in evaluation are the College of the University of Chicago, Michigan State College, and Antioch College. The Educational Testing Service, on the other hand, has been active in the evaluation of new curricula and non-traditional programs.

For over half a century now, serious consideration has been given, on a national basis as well as by individual organizations, to curriculum evaluation and to the recognition and appraisal of teaching. In addition, toward the middle of the current century, a spate of projects and programs of evaluation developed out of a concern for general education and as a reaction against overspecialization and compartmentalization. The volume Evaluation in General Education (Dressel, 1954) includes reports of the evaluation practices employed by a representative group of colleges concerned with general education. In the special area of evaluation of teacher preparation programs, a number of studies, particularly from individual institutions and organizations, have been evident. Suggestions and recommendations have been made in the use of tests and the development of instruments. The need, however, remains for unified criteria and concerted effort.

In 1924 President Raymond Hughes of Iowa State College made the first published appraisal of graduate schools in the United States. In 1934 the American Council on Education sponsored his second rating and published the report of its Committee on Graduate Education. The third major study regarding the quality of graduate education was made by Hayward Keniston in 1957.

More recent studies reflect the dramatic changes that have since overtaken academe. In the spring of 1964, the American Council on Education initiated a further and more detailed study of graduate programs, with financial support from the National Science Foundation, the National

Institutes of Health, and the United States Office of Education; and in 1966 appeared Allan Cartter's report An Assessment of Quality in Graduate Education (1966). The book is an answer to a query raised by ACE's Commission on Plans and Objectives for Higher Education on the strengths and weaknesses of graduate schools in this country "in providing well-trained scholars for both teaching and research [Cartter, 1966, p. 1]."¹ One of the purposes of the report was improvement of the assessment of particular graduate programs in major universities in the United States and examination of available techniques of evaluation. Its main concern thus was the improvement of graduate education. Despite carping critics, Logan Wilson (1970) was able to report:

Most gratifying of all was the number of institutions reporting to us that they had used the study and its results to launch a systematic effort to improve their graduate endeavor. . . . the published findings unquestionably have spurred some universities to give more careful attention to the qualitative aspects of their graduate-level enterprise [p. ix].

As planned, in view of the changing academic scene, a follow-up survey appeared in 1970 with Roose and Anderson's companion report (1970) A Rating of Graduate Programs. Updating the Cartter report, the authors presented the principal findings of the 1969 Survey of Graduate Education, in the hope that they would be of assistance to prospective consumers of graduate education, to academic administrators, and, in a time of increased concern over the accountability of higher education, to public authorities and policy-makers. Like its predecessor, this follow-up study was conceived as an aid to administrators and faculty in the upgrading of their graduate programs.

In 1970 also appeared The Recognition and Evaluation of Teaching by Kenneth E. Eble, Director of the Project to Improve College Teaching and The Committee on Evaluation. The Project was jointly sponsored by the American Association of University Professors and the Association of American Colleges and supported by a grant from the Carnegie Corporation. The conference on April 10, 1970, in Washington, D. C., of students, faculty, and administrators from a wide range of institutions of Higher Education, which provided some of the material for this study, had agreed that "wisely-formulated, wisely-administered and wisely-used student evaluations are useful to improving teaching [Eble, 1970, pp. vii]." Eble observes:

Astin's and Lee's comment seems to be a fair criticism of current practices: "If the ultimate measure of the teacher's effectiveness is his impact on the student--a view which few educators would dispute--it is unfortunate that those sources of information most likely to yield information about this influence are least likely to be used [pp. 10-11]."

The main thrust of the publication is the significance of student evaluation in the assessment and improvement of teaching effectiveness. The author deals at some length with instruments and procedures and emphasizes the necessity to establish and maintain an evaluation program. There is relevant comment also on the impact of private student response and open evaluation. The monograph is evidence, not of a conclusion that final or exact answers have been found to the disputed questions of how effective teaching has been throughout the country, how good it should be, or whether student evaluation of courses and teachers actually results in improved teaching; it is rather an

indication of the interest and inquiry these questions can generate.

Eble concludes:

If evaluation can contribute to bringing the campus together in the common teaching-learning enterprise, if it can put some life into daily routines, if it can work specific improvements upon individuals and contribute to the general enhancement of teaching, if it can add to what we know about teaching and lead us to act on that knowledge, then it is surely worth the risks and effort it may involve [p. 49].

Concurrently with such changes in educational thought, the insistence on specialism underscored the need to measure the effectiveness of our educational programs. The need for evidence on the value of general education prompted several studies that have made lasting contributions to educational evaluation. As early as 1947 the Cooperative Study in General Education, directed by Ralph Tyler and sponsored by the American Council on Education, developed useful instruments for the assessment of growth in regard to general education objectives (Cooperation of General Education, 1947). Reference has already been made in the initial section of this monograph to the Eight-Year Study of the Progressive Education Association (Smith & Tyler, et al., 1942). Although at the high school level, the objectives which this study attempted to evaluate are identical with many of the avowed objectives of general education.

Among other projects sponsored by ACE is also George Angell's survey of evaluation practices conducted in 1948. Relating these practices to general objectives it seemed, among the colleges surveyed, "that most evaluation was strictly content-oriented, but that objectives

involving more than knowledge were commonly accepted even though student achievement relative to them was not being evaluated [Dressel & Mayhew, 1954, p. 10]."

In 1954 the Council published General Education: Explorations in Evaluation, a report of the Cooperative Study of Evaluation in General Education. It was initiated in the spring of 1950 and made possible with assistance from the Carnegie Corporation of New York and annual contributions from the 19 participating institutions. The prefatory remarks of the authors conclude:

The mere fact that so many individuals would attend so assiduously to a study of evaluation in general education is one proof of the strength of the movement. Few general education teachers are satisfied with what they are presently doing and they are constantly seeking for new and better ways. We have hopes for significant developments in the future and we would like to believe that this report by its realistic summary of present practice, may serve in some slight measure to expedite such developments [Dressel & Mayhew, 1954, p. x].

With the accent on quality and effectiveness, the professional preparation of teachers and administrators is a vexed question in higher education today. The quality of college and university teaching, the influence of graduate schools on prospective teachers, and the problems of graduate practice are matters that have long engaged the attention of leaders in the field of higher education. In 1945 Ernest V. Hollis prepared for ACE's Commission on Teacher Education his report titled Toward Improving Ph.D. Programs, which describes and analyzes the Commission's activities in education and analyzes "a number of studies that served to throw considerable new light on the situation and to

delineate the issues more sharply [Bigelow, 1945, p. v]." The studies employed departmental rating procedures, student evaluation of teaching effectiveness, letters from employing groups, and follow-up of former graduates. In the three studies the inquiry highlights, the respondents were established college teachers. They were well pleased with their graduate experience, as far as intellectual values are concerned. Dissatisfaction, if any, stemmed from the quality and integrity of the professor's scholarship rather than from the educational philosophy. Most of the group approved the research techniques of the doctoral program and agreed on "the educational significance of firsthand, practical experience at some stage during a college teacher's development [Hollis, 1945, p. 171]."

The author's comment on the negative aspect of his findings offers a valuable hint on instrumentation in educational evaluation.

Despite the pronounced majority view in each set of questionnaires, it is likewise important to stress the irreconcilable minority found in each study. The nature of the instruments may well account in large measure--but possibly not entirely--for the fact that the critics were very much better at denouncing what they didn't like than in making suggestions for improvement. The point of greatest tension within each group, attested to by the vigor and emotionalism of the statements concerned, had to do with the functions and offerings of departments of education. The heaviest criticism of all tended to come to a head over courses in education and supervised practice teaching [Hollis, 1945, p. 171].

More recent studies demonstrate the growing interest in evaluation activity and new directions in measurement and evaluation. Several projects have been conceived and implemented over the recent past by

individual institutions. In October 1968 Robert T. Alciatore and Ruth E. Eckert produced Minnesota Ph.D.s Evaluate Their Training: A Study of the Relationship of Various Ph.D. Programs to Later Career Service and Satisfaction. The University's varied program offerings, its specific focus on college teaching, and its pioneering efforts and continued leadership in institutional self-study made the University an appropriate center for the conduct of this study. The inquiry sought data from a population of over 1,700 respondents on postdoctoral achievements and on the graduate experience in retrospect. Particular attention was focused on the subsequent careers and reactions of Minnesota's 33 interns who participated in the Ford Fund College Teaching Internship Program between September 1953 and June 1956. The study was designed to identify the type of skill and knowledge which recipients of the doctoral degree applied on the job and, in view of the applicability of these skills and knowledges, to appraise the doctoral program in which they were acquired. In consideration of the fact that most persons who obtain the doctorate join college and university faculties, the study devoted particular attention to teacher education. Candidates who had received formal teacher training were compared to those not similarly oriented.

The study classifies graduates in accordance with five types of preparation for college teaching and offers statistics of teaching experience prior to graduation. In the matter of publications and scholarly activities, the findings reveal that candidates with courses in Higher Education and teaching experience excel in thesis publication, while in the publication of books other than the thesis, the interns lead.

Of the 20 factors listed as possible influences in the choice of teaching as a career, intellectual challenge ranked highest.

An attempt was made also to discover motives for leaving or not entering the teaching profession, to discern factors contributory to success and satisfaction in teaching, and to explore the question of identification with teaching as a career.

The study also investigates the acquisition of skills essential to teaching and items related to administrative abilities and professional relationships. Views on controversial issues in graduate education are recorded and the authors comment:

Probably the most striking finding emerging from this analysis was the high endorsement given the proposal that doctoral candidates expecting to join college or university faculties receive more training for teaching while in the graduate school. This ranked second in the entire list, with 61 per cent endorsing it, another 18 per cent expressing uncertainty, and 20 per cent opposed to such a plan [Alciatore & Eckert, 1968, p. 54].

Increased attention was also recommended to the goals of general education and the development of administrative and teaching skills.

Chapter IV of the study provides further details of the Minnesota Internship Program for College Teachers, financed by the Ford Foundation. The inquiry then assumes a two-dimensional aspect, with the ex-interns viewing their experience from two angles, namely, impressions at the time of training and views in retrospect. The group expressed general satisfaction with the experience, but stressed the

importance of understudying a master teacher and the need for supervised teaching.

The final chapter records major findings and implications. The graduates' appraisals of the Ph.D. experience and the interns' evaluation of their special program are separately summed up. Based on their conclusions, the authors offer recommendations and suggestions for further study.

In 1969 Granville B. Johnson published a further study entitled Evaluation of the University of South Alabama College of Education Teacher Training Program by Analyses of Its Alumni, 1969. The inquiry, presented in three sections, is two-dimensional and spans the five-year period 1965 to pre-spring 1969. Section I deals with evaluation by a census population of alumni who graduated from the South Alabama College of Education during the period under review. Section II ascertains the opinions of principals on the Bachelor of Science and Master of Arts alumni of the aforesaid College teaching in the Mobile and Baldwin counties. Section III presents a summary of the study and its results, together with conclusions, limitations, and recommendations.

The general purpose of the investigation was evaluation of the alumni of the College of Education, University of South Alabama, Mobile, with a view to improving selection and training techniques at both the Bachelor's and Master's degree levels. Two instruments, severally addressed to two groups of respondents, in the ultimate analysis, converge on an appraisal of the work of the University in the development of the

skills, insights, and knowledge essential to effective teaching. The selection of the two sources of feedback on South Alabama's teacher training program is apt and judicious.

The reactions of the alumni to their respective teaching positions and the significance of their preparation for the job are described under a series of relevant heads and various tables provide details at a glance. The evaluation of principals was elicited through a rating scale on 16 variables, the first six presumed significant to teaching, and the next ten directly associated with and descriptive of actual classroom performance. In addition, "unsolicited comments" from both groups offer interesting and meaningful firsthand information.

Section III sums up the survey. The results of the two surveys, separately listed, involve description of alumni reactions and analysis of principals' ratings. The researcher states that, at the time, correlation between the data generated by the two instruments was under investigation (p. 39) and that a factor analysis was being made of these two instruments (p. 42).

A related document (ERIC #ED 060 050), by Granville B. Johnson, appeared in 1971, captioned Evaluation of the University of South Alabama College of Education Teacher Training Program by Analyses of its Alumni, II - 1971. A second major purpose was added, viz. to discover possible directional trends by a comparison of the results of the 1969 and 1971 studies. Presented in three sections again, the investigation generally follows the format of its predecessor. The design again is

that of a dual survey. The first sample comprises all alumni (B.S. and M.S.) who graduated between June 1969 and August 1970 from the College of Education, University of Alabama, Mobile. The second sample, as previously, was limited to B.S. and M.A. alumni teaching in the Mobile and Baldwin counties, whose qualifications as teachers their principals were called upon to rate. The first phase of the inquiry, accordingly, presents a self-study by alumni and the second develops as an appraisal of the alumni by their principals. The investigator sets up the results of principals' ratings:

In general, the women are rated higher than the men. This was apparent particularly in "English usage," "general vigor and enthusiasm," "general cultural background," and on a wide margin in "dependability." The differences, however, though in the same direction, are not as great as those found in the 1969 study [p. 34].

Johnson considers the 1971 study technically superior to the one completed in 1969. Begun earlier in the year, it allowed for greater initial response and follow-up probe. A doubled sample size, moreover, increased its validity and reliability.

The results of the inquiry reflect the accent on research and a greater emphasis on the need for transfer from the theoretical to practical. They reveal, at the same time, that certain specific problems reported in the 1969 study still persisted. Johnson concludes:

it may be stated that a study of this type is meant to provide feedback regarding the educational experiences of undergraduate and graduate students in the College of Education. Evaluation of the results of this feedback must result in modification both qualitative and quantitative of the experiences provided students. If this feedback is not examined

and used as a criterion for curriculum change and teaching emphasis, the purposes for performing this investigation have not been met [p. 47].

In January 1972 C. Jarvis Wotring completed A Pilot Study to Evaluate Teachers Educated at Central Michigan University. The results of this important study are presented in section four, Activity 1-2, page 58 of this paper.

The new ideas in education have prompted studies that reflect the new thinking and represent departures from the traditional. In August 1971 Lee Napier published Implementation of a Competency-Based Teacher Education Program, Final Report, Volume I. The purpose of the project was to assist implementation of the change from traditional instruction programs in teacher education at Jackson State College, Mississippi, to programs that are competency-based. This initial attempt focused on the faculty and interns of the Jackson State College--Hinds County Teacher Corps Program.

The researcher provides a relevant description of the setting. Thirty interns (originally 36), all Blacks, and taking their college course work at Jackson State College, were placed in three public schools in Hinds County, Mississippi, namely, the Bolton Attendance Center, Lovette Elementary School, and Westside Junior High School. The University of Toledo Teacher Education Model was selected for the competencies and format to be used. During the period July 1970 to July 1971, the interns were given eight courses in Early Childhood Education and six hours of internship that would count toward a

Master's degree in Early Childhood Education. Each course constituted a module in competency-based instruction to some extent, varying from 50 to 100 percent.

Student evaluation of the program revealed positive reactions to competency-based instruction, in that it involved less pressure and brought higher achievement and higher grades. The instruments used were questionnaires twice during the year. The Teacher Corps faculty were also required to appraise the program as far as it concerned their particular involvement in it. Their reactions were equally favorable. Chapter IV lists the conclusions reached and the changes that conversion from traditional instruction to competency-based instruction will necessitate in the present institutional structure.

The concept of performance-based teacher education has received equal emphasis. In January 1971 the Division of Elementary and Secondary Education, Florida State Department of Education, Tallahassee, published a study captioned Individualized Inservice Teacher Education: A Performance-Based Module. Originally designed for use in a 75-minute workshop session, the module is adaptable for use in other situations. The document includes an evaluation sheet, a post-assessment achievement test, a key for evaluation, practice exercises, and a component checklist. In December of the same year appeared a related document (ERIC #ED 058 166), by Stanley Elam, American Association of Colleges for Teacher Education, titled Performance Based Education: What Is the State of the Art? and sponsored by the United States Office of Education. The document clarifies the concepts of performance-based

teacher education, examines its potential, and identifies advantages and the attendant problems and issues.

Among individual projects in the evaluation of teacher training programs may be mentioned Clinton A. Erb's A Formative Evaluation of An Experimental Teacher Education Project for Juniors in Mathematics Education, a paper presented at the Fiftieth National Council of Teachers of Mathematics Education. Heath (1969) writes:

A series of evaluative studies of new science and mathematics curricula were reported as these courses were introduced on a national basis (Maier, 1962; Payette, 1961; Stickell, 1965; Heath & Stickell, 1963; Grobman, 1964; Rosenbloom, 1962; Shuff, 1962). These studies, for the most part, employed conventional achievement tests. Measures of cognitive preference or problem-solving style have been the object of later research (Heath, 1964b; Travers, 1965) [p. 282].

The Open University is another new concept in education that has provoked evaluative study. A Summary of the Instructional System Used by the Open University in Great Britain in 1971, by David G. Hawkrigde (1971), describes the characteristics of the system and the major learning resources it offers students. The document describes how course units have been developed and provides details of the steps being planned to evaluate the effectiveness of the system.

Reference, however, has already been made to a more recent study on the adaption of the Open University concept in American institutions, directed by Rodney T. Hartnett, Educational Testing Service. The study was conducted in cooperation with the College Entrance Examination Board and financed with a grant from the Carnegie Corporation. The Chronicle

of Higher Education, dated May 28, 1974, quotes from a report of the study:

"The Open University courses required a good deal of time and effort, and rather traditional student competencies. As such, these courses probably should not be regarded as a particularly promising means of extending a postsecondary educational opportunity to those who might be regarded as educationally disadvantaged."

.....
 "The courses clearly appealed most to those who had already demonstrated academic competence, came disproportionately from high-status occupations, and reported relatively high incomes [pp. 1-2]."

A number of other institutions, too, have conducted studies in the evaluation of teacher preparation programs. As individual reference to these studies is beyond the scope of this monograph, they have been included in the Directory of Institutions appended to this work.

Present practice in the evaluation of teacher preparation programs thus reveals some effort by individual institutions and organizations. The time, however, demands a more concerted and sustained drive, in view of which the authors of this paper would reiterate their plea for joint institutional effort by institutions of teacher education and for communication and cooperation between these institutions and the school systems. In 1954 Dressel and Mayhew wrote:

Evaluation of educational achievement is so complex a field that even test experts are humble in the presence of the many complications and constantly seek to improve their understanding as well as their techniques. The need to measure as surely as we can the efforts of our educational programs is a compulsion felt by all educators and administrators. Any new light on this problem is therefore a stimulus and a challenge [p. vii].

Applied in the special context of teacher education, this opinion, two decades later, still warrants confirmation.

SECTION IV

A MODEL FOR ASSESSMENT OF TEACHER PREPARATION PROGRAMS

Introduction

As the preceding sections have shown, it is not only important from a practical standpoint to assess the effectiveness of your teacher preparation programs, but also absolutely essential if teacher preparation programs are to remain viable. Higher education has a responsibility to become increasingly more responsive and accountable to those it serves. Increasing pressure will be put on schools of education to not only show that they are doing a "good" job, but also, in some cases, to justify their very existence. The question, then, is not whether to assess teacher preparation programs but how do we do it?

We recognize that the best indicator of a successful teacher preparation program is the performance of the product, which in this case is the teacher, and subsequently the performance of the learners in the instructional system of which the teacher is a part. In other words, how well do the learners perform in society and does their behavior contribute to the good of the society of which they are a part? What we are specifically referring to is the growing conviction that teacher performance should be tied to the concept of "student-gain." Has the learner achieved the intended objectives of the instructional system of which the teacher is a part?

Although worthy of consideration, these types of measures are extremely elusive. The state of the art, at the present time, leaves us in a state-of-limbo if we are to attempt to use the performance in society of a teacher's students as the final criterion in determining the effectiveness of our teacher preparation programs. Although we are not saying it is impossible, we are saying that at the present time we do not know how to effect this type of evaluation program at a reasonable cost and a reasonable turn-around time in obtaining and processing the data. In addition, what data do we collect? The amount of money a person makes? His position in the social ladder? The type of job he holds? His contribution to society? There are so many factors that must be considered that even if we had the resources it would involve an expenditure of time that is beyond our capability to expend.

Therefore, what we offer, in this working paper, is an intermediate step in evaluating the quality of your teacher preparation programs. The model proposed here is not perfect and must be considered in the light in which it is presented--a model that can be used by your institution as a guide in establishing your own individualized programs of evaluation. In most cases the model presented here will have to be modified by your institution to fit your specific needs. What is presented here is a model to guide you in setting up a more complete model for your specific goals.

We recommend that each institution set up what might be called a Teacher Preparation Assessment Group (TPAG). This group would be

responsible for the preparation and implementation of the assessment instruments; the analysis and interpretation of the data; and the constant upgrading of the instruments, procedures and analysis. It is further recommended that the TPAG be made a permanent part of the institutional structure. It is only in this type of structure that the group can be effective in accomplishing the goal of the assessment --to continually provide information to both the public and the institution about the quality of its teacher preparation program. An institution armed with this process of dynamic evaluation procedure can only enhance its position in terms of the quality of its program and, hence, its credibility.

Since your program is constantly undergoing an evaluation process you will, in effect, be taking steps to become accountable for your program. We believe this to be important because it exemplifies a leadership function that is all too rare in our current teacher preparation programs. This leadership goes far beyond reacting to the pressures of the times--it is the process of creating new structures/procedures to cope with the changing demands of a rapidly expanding technological society.

Assessment Procedures

Among many alternative procedures for evaluating the quality of your graduates, the two methods that seem to be mentioned most often in the literature are interaction analysis and survey analysis. Although other procedures are available such as case studies and interview analysis, these procedures are much too expensive to conduct and maintain.

Therefore, we will address ourselves to these two most common procedures.

Interaction Analysis

Interaction analysis is a technique to establish the degree of a teacher's control of the student's freedom of action. The teacher's actions are classified into two categories: those actions that inhibit the learner's freedom of action and those actions that increase the learner's freedom of action. But, interaction analysis has severe limitations in its implementation and usefulness as an effective tool in assessing teacher preparation programs.

In terms of implementation it is a system that requires an observation in the classroom of the individual teachers. This in itself makes it economically impossible if you are interested in assessing the effectiveness of a teacher preparation program over a long period of time. Although implementing it for a one-shot evaluation of a small sample of teachers should not be beyond the resources of the average college of education, it is economically impossible to use it for any large scale evaluation or one that requires a dynamic component.

In terms of its usefulness current methods of interaction analysis deal only with verbal behavior or the lack of verbal behavior in the classroom. This preoccupation with verbal behavior severely limits interaction analysis as an indicator of the effectiveness of the instructional program, the teaching, and, consequently, the learning that is supposedly taking place in the classroom.

The Flanders' System of Interaction Analysis, a relatively simple system to operate, is the most popular. A great deal of work has been done with Flanders' System and it has been shown to be both valid and reliable. The Flanders' System has only ten categories: two for learner talk, seven for teacher talk, and one for silence or confusion (Simon, 1967).

Other systems such as the Verbal Interaction Category System (VICS) by Edmund Amidon and Elizabeth Hunter (1966) and the Hough System by John B. Hough (1965) are both extensions of the Flanders' System. Amidon and Hunter have expanded Flanders' System to include five categories for analyzing classroom verbal behavior: teacher-initiated talk, pupil-initiated talk, teacher response, pupil response, and other which includes silence and confusion. Hough has expanded Flanders' System to 16 categories and includes some non-verbal behaviors as relevant activities.

As was mentioned before, all of these systems require direct classroom observation and are highly dependent upon the use of verbal behavior as an indicator of learning. In many cases, though, verbal behavior cannot and should not be used as an indicator of student learning. One example should suffice: the typical science laboratory class. In this type of instruction setting, there may be silence and observational confusion which would be judged in a negative manner. The resulting conclusion would be that learning was not taking place when, in fact, a great deal of learning was occurring. Other instances would be open classrooms and self-paced, self instructional packages.

Interaction analysis cannot be used in these situations. Although we recognize the fact that learning takes place through verbal interaction, we must also recognize the fact that learning can and does take place in the absence of verbal interaction. Therefore, for these reasons plus cost and logistical considerations, we do not feel that interaction analysis is the answer to the dynamic evaluation process of teacher preparation programs.

We would like to emphasize that we are not precluding the use of interaction analysis as a tool of evaluating teacher preparation programs. It is extremely useful in limited situations and can be used as an adjunct to other procedures. For those who are interested in using interaction analysis as a procedure, we would refer the reader to J. T. Sandefur's An Illustrated Model for the Evaluation of Teacher Education Graduates published by the American Association of Colleges for Teacher Education (see Reference Section: Sandefur, 1970).

Survey Analysis

This procedure appears to offer the most for the money. Survey analysis can provide large amounts of data at a relatively small cost. It can also be used to query a larger sample than would otherwise be feasible. It also is a procedure that lends itself to modification and can be used as a dynamic evaluation procedure. That is, a systematized procedure can be initiated to provide your institution with a continuous method of evaluating your graduates and, consequently, your program through a continuous feedback loop. It is estimated that at the end of a three year period an institution would be able to effectively determine

whether its teacher preparation program was meeting the needs of the learners or not and to build into its program a constant updating of its methods of teacher preparation based upon a continuous cycle of feedback from its graduates.

Although survey analysis is not the perfect method nor the final answer in evaluating teacher preparation programs, it does offer at the present time the most cost-effective method of obtaining data. Therefore, our model is based on the use of survey analysis as the primary means of assessing the worth of any particular teacher preparation program. As stated before, there are other procedures that can be used in addition to survey analysis and they should be used if it is feasible.

In the generation of the model we have assumed the ideal situation which is one of unlimited resources. We recognize this is not the case in the real world and compromises must be made. Therefore, the model presented here is one of generality and is presented to serve as a guide. It is anticipated that your institution will develop its own modification of the model and, hopefully, improve on its usefulness and effectiveness. "It is a bad plan that admits to no modification" (Publius Syrus, 100 B.C., Maxim 469).

The purpose of conducting a survey analysis is to answer certain questions that will allow you to make decisions about the quality of your teacher preparation program. It is the determination of these questions and, consequently, the specification of the objectives of the study that is an extremely important activity. The successful completion of this activity depends, to a great extent, on the

understanding of the general functions of survey analysis. Therefore, before describing the specific activities of the model it will be useful to examine certain general aspects of survey analysis.

Functions of a survey. Surveys serve two general functions: (1) to collect data for the purpose of making statements about certain attributes or traits of the population of graduates, based solely on descriptive types of data; and (2) explaining why these certain attributes or traits occur.

The first function is limited to the collection of data for the purpose of making descriptive statements about the sample of graduates. You, the researcher, are concerned only with describing certain traits or attributes of the sample. You are not concerned with the reasons as to why these traits or attributes occur, but only in describing their distribution. If the sample of graduates is carefully selected (this will be discussed in the activity entitled "Selection of Sample"), the descriptions of this sample can lead one to safely infer the characteristics of the population.

In addition to describing the total sample, it is also possible to make descriptive statements about certain subsamples that may exist within the total sample. For example, if your total sample represents all types of teacher graduates an appropriate subsample would be those teachers trained to be elementary school teachers. It would then be possible to make descriptive statements about this specific subsample. This would be helpful if you want to assess the programs of the separate departments of your school.

The second function is to provide an explanation of why the observed distribution of traits and attributes occurs. This, in most cases, involves the use of fairly sophisticated statistical procedures. In this case we are interested in discovering relationships between two or more variables and inferring why the distribution exists, in addition to describing the nature of the distribution.

It should be noted that a wise TPAG either includes a statistician or consults one at this point to assess the proper statistical procedures that will be necessary to answer the objectives/questions of the survey. Many studies fall apart at the end because of inadequate statistical preparation at the beginning of the study.

Unit-of-analysis. One last general aspect of survey analysis should be touched upon before proceeding to the descriptions of the activities listed in the model. This is the concept of the unit-of-analysis. In your particular case of teacher preparation programs the typical unit-of-analysis is a person, or more specifically the individual graduate of a program.

However, there is no reason why this must be the case. One might consider department or teaching field to be the unit-of-analysis and, in many cases, this is all that is necessary. But if department is chosen as the unit-of-analysis there will be no way to "go back" and look at individual graduates after you have collected the data concerning departments. Therefore, it is wise to choose the smallest unit-of-analysis if practical (and it is in our case) because the use

of person/graduate as the unit-of-analysis does not preclude the use of larger units such as departments later on in the study. What is important is deciding in advance the proper unit-of-analysis for your study. If this is not done, the sample design and data collection may not allow for an appropriate analysis.

Major Activity Sequence

The following presents a Major Activity Sequence of the activities to be completed in your survey analysis assessment of teacher preparation programs. The Major Activity Sequence Chart is presented on page 57. A description of each activity is provided in the remainder of Section IV. Since some of the activities of the survey analysis procedure occur concurrently with other activities, it is not possible to present a description of the activities in a chronological fashion. Therefore, each activity is described in a separate portion of the text with the activity number listed in the upper right hand corner of the beginning page of that activity.

It is not necessary to read the activities in the order presented. If you want information regarding the selection of the proper sample size, then you only need to refer to the portion of the text labelled ACTIVITY 9-12: DETERMINING SAMPLE SIZE. The activities are listed for your convenience on the following page with their page numbers. However, you should consult the Major Activity Sequence Chart to determine the proper sequence of the activities and their relationship to the other activities.

	<u>Page</u>
ACTIVITY 1-2: FORMULATION OF PROJECT AND OBJECTIVES	58
ACTIVITY 1-15: DOCUMENTATION	63
ACTIVITY 2-4: SELECTION OF SURVEY DESIGN	65
ACTIVITY 4-6: ITEM CONSTRUCTION	71
ACTIVITY 6-7: INSTRUMENT FORMAT AND CODING	78
ACTIVITY 2-3: SELECTION OF SAMPLE	85
ACTIVITY 3-5: SELECTION OF SAMPLING FRAME	89
ACTIVITY 5-9: DESIGN OF SAMPLING PROCEDURES	92
ACTIVITY 9-12: DETERMINING SAMPLE SIZE	95
ACTIVITY 7-10: DEVELOP PROTOTYPE INSTRUMENT	100
ACTIVITY 10-11: PILOT TEST AND REVISION	103
ACTIVITY 11-12: DEVELOPMENT OF FINAL INSTRUMENT	105
ACTIVITY 4-8: SELECTING STATISTICAL ANALYSIS AND PROCESSING OF DATA	108
ACTIVITY 8-12: PROCEDURES FOR PROCESSING RETURNS	114
ACTIVITY 12-13: INITIAL MAILING	117
ACTIVITY 13-14: KEEPING TRACK OF RETURNS--CHARTING AND FOLLOW-UP	118
ACTIVITY 14-15: METHODS FOR ANALYZING RETURNS	122
ACTIVITY 15-16: PREPARATION OF FINAL REPORT	125

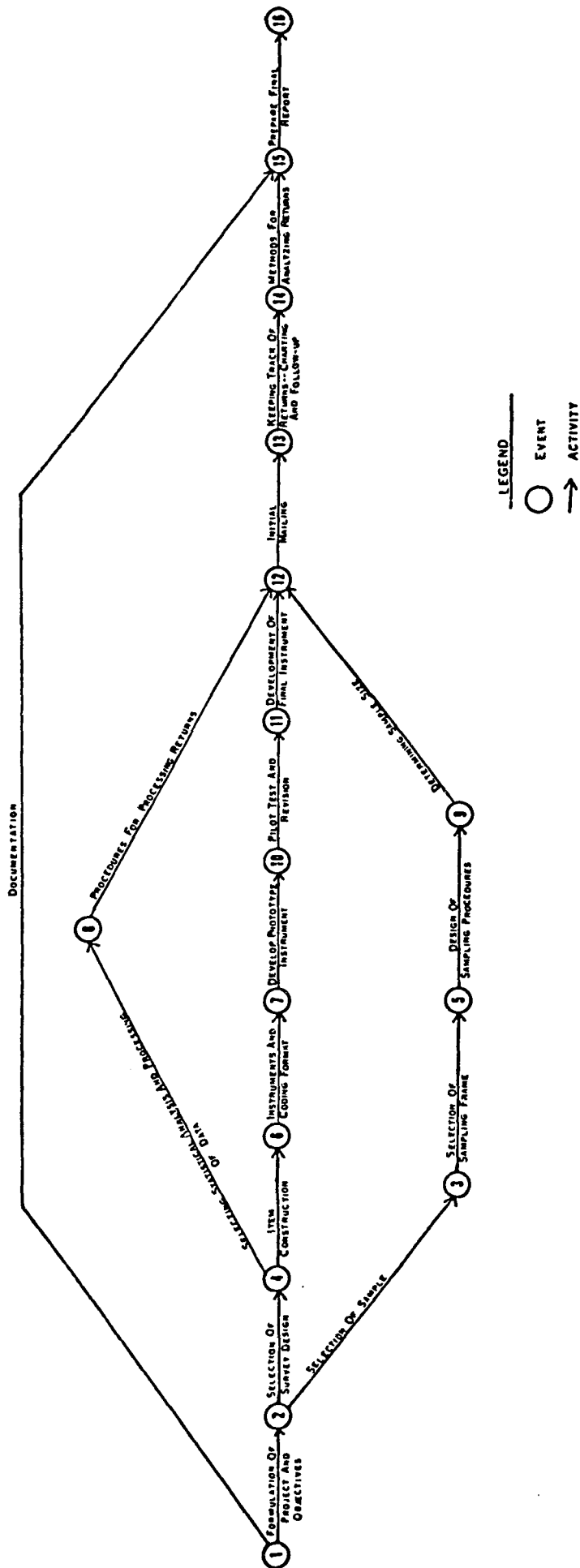


Figure 1.--MAJOR ACTIVITY SEQUENCE SURVEY ANALYSIS
ASSESSMENT OF TEACHER PREPARATION PROGRAMS

ACTIVITY 1-2FORMULATION OF PROJECT AND OBJECTIVES

Is there a need for assessing the program of the college? In most cases the answer will be a resounding "YES" and the evaluation procedures will continue. If, however, after due deliberation you are satisfied with your present teacher preparation program and are confident that you are producing the best teacher possible, then you obviously do not have a need to continue in this endeavor and should terminate all related activities. If this be the case, then please write to us and tell us how you produce the perfect teacher. All the world awaits.

Since we are confident that few people will be writing to us, we will proceed with the immediate activity at hand--which is to define the objectives of the study. This is certainly one of the most difficult tasks of the entire procedure. However, it cannot be overemphasized that the determination of the objectives is of extreme importance because every activity and, in fact, the results of the study depend on the precise definition of the assessment objectives. It cannot be left to chance.

You are interested in both the immediate effects of your training program and the changes that take place over time to the sample. Can you deduce these changes as a result of your training program or are they due to external influences? In either case you want to explain why and what factors could be causing the changes. This emphasizes

the need for a type of dynamic evaluation procedure. That is, you must at regular periods assess your graduates in terms of their training and job performance and to use these results to upgrade your program. Formulating the goals and objectives of the study must take this process into account.

Although we cannot define the objective of your particular survey analysis, it is possible to list several options open to you in determining just what should be done.

- (1) You can assess only the program of the College as a whole.
- (2) Assess only the individual programs of the departments or organizational units within the College.
- (3) Assess selected departments. For example, you might be interested, at this point in time, in the performance of your Secondary Education effort since this department is responsible for the largest number of graduates and will have the greatest impact on the outside community.
- (4) Assess one of your smaller departments and use this department as a pilot project for refining your evaluation methods. This will provide a great deal of information for later use in the assessment of your total program and other departmental programs. Although you might expect to save money by taking this route, the savings may be illusory since one must expect certain minimum costs regardless of the number of departments or graduates involved. However, if the final result of your assessment is concerned with a

large number of graduates, then a pilot run at this stage might be advisable.

- (5) Assess both the College as a whole and the individual departments. In this case many activities will be concurrent and in many cases be duplications. There are many advantages to this route such as avoiding the cost of duplication of tasks plus having central control over the overall activities. It is for this reason that we have recommended the establishment of the TPAG. All activities can be coordinated through this group and will subsequently be cheaper than if each department were to conduct its own assessment.

A study conducted at Central Michigan University used survey analysis to evaluate their teacher preparation program in terms of specified objectives (Wotring, 1972). Examination of the objectives and partial results of this study may help you in formulating your own.

Briefly, the purpose of this study was to ask principals who had Central Michigan University graduates on their staff to evaluate the teachers in terms of eleven criteria. The same teachers were also asked how they felt about their preparation. The instrument provided for evaluation based on the following criteria: (1) Planning and Organization, (2) Methods and Materials, (3) Motivation, (4) Evaluation, (5) Management, (6) Overall Classroom Effectiveness, (7) Professionalism, (8) Community Skills, (9) Academic Preparation, (10) Personal Qualities, and (11) Human Relations.

The study answered eight questions, four of which were:

- (1) How did the teachers view their preparation?
- (2) Would teachers and principals agree on their evaluations?
- (3) Could important factors of teacher evaluations be identified?
- (4) Would grade point averages and American College Test scores predict success in teaching?

The results were (1) the teachers had a positive view of their preparation and the way the teachers felt toward their instructors affected their satisfaction with their education; (2) the teachers and principals did agree on their evaluations; (3) three factors-- (a) skill factors, (b) human factors, and (c) academic factors--should be included in the evaluation of a teacher's performance; and (4) grade point averages (GPA) and American College Test (ACT) scores did not predict success in teaching.

This study was conducted with definite goals and objectives in mind with the survey analysis instrument designed to ask questions related to these objectives. Subsequent data reduction techniques were then employed (factor analysis was used to identify the relevant skill areas) to answer the specific questions derived from the objectives.

The objectives you eventually define for your assessment will be based on many factors unique to your particular situation, such as: Do your graduates primarily teach in an urban setting? If they do,

then your objectives would be addressed to assessing your program in terms of their performance in urban schools. Of course, if you do not know the particular environmental settings where your graduates teach, then one of your objectives would be to determine the environmental setting.

There are no hard, fast guidelines to follow in deriving your objectives. They ultimately depend upon the goals of your study, the questions you want answered, and the amount of money you have available. Although we stated in the introduction to this section that we would assume the ideal situation which means unlimited resources, this is not the case. Compromises must be made and the compromises must be based upon your unique situation.

ACTIVITY 1-15**DOCUMENTATION**

This activity occurs concurrently with every other activity in the model. The TPAG should document each procedure and process as it occurs. Starting with a statement of the problem every phase should be written down. Mistakes as well as hits should be recorded. Improper procedures should be noted and logistical problems, processes and procedures should be well documented. The reason for all this documentation is twofold. The first is that when it comes time to prepare the final report the task is considerably reduced if activity notes are maintained throughout the course of the project. Secondly, the documentation allows the TPAG to correct deficiencies in procedures so that the next survey analysis can be more efficient.

The documentation will reveal overlapping of responsibilities and procedures by individuals in the projects. It will also reveal activities that were poorly conducted or totally neglected.

Although it may seem mundane for us to remind you of the necessity for documenting each of your activities in as much detail as possible, very few researchers who conduct survey analysis document their procedures. This leads to duplication of mistakes the next time the research is undertaken. This lack of attention to documentation leads not only to wastes in terms of expended man-hours, but it is economically a poor decision. Any leader that does not require his "troops"

to document their activities is negligent in his duties and should be requested to pay for any similar mistakes committed by the following group. If the documentation is taken seriously and maintained over a period of years, we believe it is self-evident that an extremely efficient TPAG could be maintained by the college at a reasonable expense. Since it is the primary responsibility of this research to evaluate the worth of your teacher preparation programs, an efficient procedure for conducting this evaluation on a regular basis is essential if your program is to cope with the every changing demands placed on it.

ACTIVITY 2-4**SELECTION OF SURVEY DESIGN**

There are two major types of survey designs to consider:

(1) cross-sectional survey, and (2) longitudinal survey. In the cross-sectional type of survey you are interested in collecting data from a sample of a specific population at one point in time. For example, if you were interested in establishing teacher attitudes toward a proposed legislative act regarding the institution of added requirements for a teaching credential, you would use the cross-sectional type of survey design because you are only interested in collecting information regarding this issue at the point in time when the legislative act is proposed.

In the longitudinal type of survey you are interested in collecting data from different samples of people from the same population over different periods of time. If the above proposed legislation has been enacted into law, you then might be interested in establishing teacher attitudes toward the law over a period of time to discover if a change of attitudes had or was taking place. In this case your population would be the same--teachers who were affected by the legislation--but your samples would vary from year to year due to the change-over of teacher personnel. Using this design allows you to compare teacher attitudes over a period of time.

There are several different types of longitudinal designs that should be considered. These are: (1) trend studies, (2) panel studies, and (3) cohort studies.

Trend studies are somewhat self-explanatory. They allow one to examine a specific population at different points in time through the use of different samples selected from the specific population. In the case of the proposed legislative act, you might be interested in determining if a change of attitude was occurring in the teacher population through the course of the discussions about the proposed legislation. You would then sample the population of teachers at selected time periods throughout the course of the discussion. These time periods could be preselected or you could conduct an investigation at key points in the discussion when major news might be generated by the discussions. In most cases, trend studies are conducted over long periods of time. Therefore, your institution must be committed to the evaluation of its teachers over a relatively long period of time if it is to detect any significant changes. This involves not only a commitment of time, but an adequate allocation of resources to the TPAG to conduct an appropriate study.

A panel study differs from other types of longitudinal studies in that data is collected from the same sample over a period of time. It is obvious that this type of survey design is not appropriate for establishing data about the quality of your teacher preparation program as a result of changes in your program.

If you were to sample 1973 graduates and ask them questions about their job satisfaction and then to sample them a year later asking the same questions, you could compare their job satisfaction from one year to the next. But this would not help in establishing the effectiveness of your training program for the general population of teachers that have since graduated from your institution.

Although the panel study is the most sophisticated of the longitudinal type of studies, it will have a tendency to become unmanageable after a period of time. Panel studies are expensive, time consuming, and suffer from a malady called panel attrition. Members of the sample move, die, change positions, and over time tend to disappear, or rather, to become unavailable. It is doubtful whether many institutions would be willing to commit the resources to conduct this type of study.

The cohort study is concerned with establishing characteristics about a specific population over time. To conduct a cohort type of study you would select a sample of the 1973 graduates and examine their attitudes toward the enacted legislation. Next year you would again sample the 1973 graduates and compare their current responses to their previous responses. You could then measure a change of their attitudes over time. This is similar to a trend study except that you are using different samples from the same population--in this case the class of 1973. A trend study would compare the responses of a sample of 1973 graduates to a sample of 1974 graduates. In the cohort study you would always be describing the class of 1973, whereas in the trend study you would be describing the population of all classes of graduates.

Each of these types of designs offer the researcher a unique set of data with which to work. Which design you use, again, depends upon your unique set of objectives. If you are interested in assessing your program in terms of graduates within a specific year without making comparisons between years, then the cross-sectional design is an appropriate and economical design.

However, we feel that if it is the desire of an institution to evaluate its program in terms of its graduates, it should be interested in improving its program over time. To accomplish this task, data must be collected at periodic intervals to assess the overall worth of the program. This data should be in a form that is generalizable to the entire population of graduates. Different samples over time can then be contrasted to determine if programmatic changes have had any significant effect on the performance of your graduates.

In addition, you need feedback from your graduates to inform you of any needed changes over time. To accomplish this a dynamic evaluation procedure must be initiated and implemented over a relatively long period of time. Therefore, some form of a longitudinal study must be set up. It isn't necessary to stick to one particular design; combinations and variations of the basic design can be used to fit your circumstances.

A variation that seems to us most appropriate is the use of dual samples. In this case you are interested in comparing the responses of one or more samples from one or more populations. We

recommend to the reader that your survey design be set up to accommodate both graduates and their immediate supervisors. If you query both samples, you can then compare their responses to see if there is agreement or disagreement. For example, you might ask your graduates to rate their performance as teachers using some sort of rating scale (see Activity 4-6). At the same time you ask their immediate supervisors to rate the teacher's performance according to the same scale and then compare their responses. Here you are comparing two samples from two different populations. More appropriately, you might ask the graduates to evaluate their training program in terms of their present teaching performance and ask the supervisor to evaluate the teacher's performance in terms of the teacher's training. Again you can compare their responses for agreement or disagreement. In this particular case of dual sampling we are really using the sample from one population to generate the sample from another population. This is perfectly acceptable and in this case necessary if you are to match the responses of the graduates to the responses of their immediate supervisors.

In summary then, each survey design is used to collect data from a specific sample of a given population in order to provide information to describe and explain specified characteristics of the given population which they represent. You then can compare samples within a general population, samples from different populations or samples within a specific population. In all cases you can describe changes that may occur over time or as a result of different training procedures.

This description of Activity 2-4 is certainly not exhaustive in its content regarding the different types of designs available. No one way is correct and we must again dodge giving an absolute answer by referring the TPAG to their stated objectives as guidance in choosing an appropriate design. There are many fine books available on the subject and the TPAG should investigate thoroughly the many different ways to set up their survey design before proceeding to the next activity.

ACTIVITY 4-6
ITEM CONSTRUCTION

The purpose of this activity is the development of an item set for the prespecified objectives of this study. We will provide some basic guidelines for both essay and forced response items. The essay type item simply requires the graduate to construct his response. The forced response mode provides a set of responses from which the graduate selects one or more. Following is a section which deals with the development and comparative advantages of the two approaches.

Forced vs. Essay Response Mode

Selection of an appropriate response mode is crucial to the resolution of the hypotheses (i.e., objectives) of the study. Also, this activity dictates procedures for the subsequent processing of the data.

The use of the essay response mode (free response) allows for a great deal of latitude in the interpretation of items by the graduate. This is both an advantage and a disadvantage. It is advantageous because the instrument is very sensitive to the individual feelings of the graduates. It also allows the graduate to provide information which is not directly requested. For example, a question may deal generally with the adequacy of a course in a particular skill area. Essay formatting many times will elicit not only evaluation of the skill area, but also reference to units of instruction, other courses, faculty behavior,

etc. This information may be helpful for recommendations and alterations of curriculum.

A rigorous analysis of items paired to objectives must be completed and the questions should be written to direct the graduate to respond to the specific area requested. These steps may seem somewhat tedious but are necessary to the success of the study. When this procedure for developing an essay type format is followed, a viable instrument will result.

Global responses such as those elicited in essay mode questions do have negative aspects. The primary problem involves quantifying such responses when they are so divergent in nature. Many times when global responses are requested, the students' attitudes are hidden in a maze of superficial comments or, in many cases, the issue presented by the question is not even answered.

The essay response mode has many drawbacks in terms of efficiency, interpretation, and analysis. The term efficiency in this context refers to both the behavioral and cost outcomes when this type of instrument is implemented. Because of the magnitude of the work involved in filling out essay questionnaires, the return rate will usually be relatively low. This dictates that additional instruments must be prepared for subsequent mailing.

The interpretation of the responses on essay questionnaires may cause a bias in the data. In other words, the essay responses must be interpreted and quantified by your TPAG. This may become costly in

terms of the man hours and in terms of the accuracy of the interpretations. Effective analysis requires that a classification of responses be developed for each one of the items in the questionnaire. This quantification is complicated by ambiguous responses and inconsistencies between raters (i.e., interrater reliability).

Analysis of essay items is a twofold problem. First, the qualitative responses must be reviewed and interpreted into a quantitative format. Secondly, the interpreted data must then be analyzed with consideration to the specified objectives. Although the essay format is advantageous because it provides additional information by allowing graduates to freely comment on areas of interest, the shortcomings in terms of interpretation, quantification and costs are of such a magnitude that we do not recommend essay items except for very small studies and/or fact-finding missions.

The forced response or multiple selection approach has four main advantages:

1. It allows the use of specific closed-end questions. This provides a means of matching the objectives to sets of concise questions.
2. It requires less effort on the part of the graduates.
3. The necessity of interpretation of responses can be eliminated prior to analysis since the responses are closed-end.
4. The encoding of information into a computer compatible format is relatively simple.

The disadvantages of the forced response method are primarily in the area of the derivation of the item set. Since the questions elicit a quantitative response from the user, care must be taken to develop items which cannot be misinterpreted. Since all aspects of prespecified objectives must be included in the item sets, the graduate cannot freely comment. Another disadvantage occurs when objectives do not lend themselves to closed-end questions. Even though various aspects of an objective are tested through an item set, the concept of the objective may not be amenable to this format.

The issue of response mode involves the review of the objectives of the study and the evaluation of the resources available. The process of the derivation of item sets for the objectives of the study is somewhat the same regardless of response mode. The primary purpose is to provide a set of items which adequately cover the objectives. Concurrent with this purpose is the development of items which are unbiased in their questioning of the graduates. For example, questions stated in either a positive or negative manner may lead to a response which would not be elicited if the question were posed objectively. A secondary purpose is to provide a set of items which maintain a thread of consistency throughout the questionnaire. A conglomeration of items which do not follow a consistent format requires additional instructions to the graduate and may lead to confusion and frustration on his or her part.

Attitude Scales

There are three types of attitude scales in common usage today. They are: (a) summated rating scale, (b) interval scales, and (c) cumulative scales.

Summated rating scale. The most common type of summated rating scale is called the Likert-type scale. The Likert-type scale requires the graduate to choose one of three or usually five categories which indicate his attitude toward the presented statement. The response categories are ordinal in nature. The most common set of response categories are:

Strongly agree
Agree
Undecided
Disagree
Strongly disagree

The Likert scale can be used in many different types of situations. The response categories do not have to be the ones listed above, but can be any set that represents ordinality. The Likert scale is very popular because it is easy to construct, simple to score, and replaces complexity of interpretation with simplicity. This type of scale is called a summated rating scale because the response to each item can be given numerical values 1 - 5 and summed across all items.

Equal interval scales. Equal interval scales require the assignment of values to the specified responses regarding a certain variable. This assignment of values is accomplished by having a group of "experts" assign values, say from 1 to 10, to the responses in terms of how intense the statement relates to the variable. If the variable to be

considered is "The excellency of teaching," the experts would be asked to assign the highest value to the statements which were a strong indication of excellency-of-teaching and the lowest value to the statements which were weak indications of excellency-of-teaching. After the experts had completed their tasks, you would then select those statements which produced the greatest agreement between the "experts." The statements would then be ordered from 1 to 10. The statements that did not produce agreement between the experts would be discarded. You would then have a set of statements relating to a particular variable with an equal interval between them. Although the concept of having a set of statements with equal intervals is highly desirable, the task of generating the statements is very complex.

The time and money involved in generating the item is, in many cases, prohibitive. In addition, the values assigned to a variable may change over time which would require a constant updating of the statements and their valuation by the experts. We do not recommend this type of scale unless your resources (time, money, and personnel) are adequate.

Cumulative scales. The cumulative type of scale is frequently called the Guttman scale. It is a scale that is unidimensional in nature. That is, if a graduate responds to a statement that is designed to indicate a strong intensity toward a variable, he is assumed to agree with those statements that indicate a lesser degree of intensity toward the same variable.

The process of generating a Guttman-type scale, like the interval scale, is a rather lengthy process of determining the attitude variable,

generating a set of descriptive statements that represent qualities of the variable, determining the degree of homogeneity of the statements, and rank ordering the statements.

Complete descriptions of the method can be found in many psychometric texts and related articles. Among these are Guttman (1944) and Guilford (1954, 1965).

Item construction is a complex task that requires time, knowledge and manpower. We again stress the importance of relating your items to your objectives. If this is done, then the choice of what type of item to use is relatively easy. If your objectives are "fuzzy," then item construction becomes a real headache which can fast turn into a nightmare. The feedback loop between item and objectives is essential if your instrument is going to do the job you want.

ACTIVITY 6-7**INSTRUMENT FORMAT AND CODING**

Format and coding of your instrument is an activity that is just as important as the derivation of the item set. Your response rate can be affected by the ease and understandability of your instrument. If your statements and questions are not arranged properly, the graduates may miss items and they may misinterpret your items leading to incorrect types of responses. If your items and instructions are not arranged in a meaningful manner, the graduates will become confused and provide you with a useless data set. Your error in disregarding the importance of formatting and coding will, in the end, not only result in a faulty data set but also may result in the disposal of the survey instrument by the graduate.

Formatting

Each section of your survey instrument should contain sufficient cues to the graduate to enable him or her to understand and successfully complete the instrument. The statements should be clear and concise. It is good practice to provide an example if the procedure is somewhat complicated or is unique in its process. If your survey instrument is arranged into subsections each subsection should contain its own specific set of instructions. You should tell the graduate the objective of the section. If the graduate encounters difficulty, he or she, in knowing the objective, may be able to solve the discrepancy.

Although stressed in another activity, we feel it sufficiently important to stress the point of spreading out your items. Do not crowd them together--give plenty of room for responses. Also make sure the graduate knows where to place the appropriate response.

Avoid lengthy items. If your items are long-winded they are likely to lead to confusion and misinterpretation. Many times this problem can be cured by evaluating your items in terms of the following two points: (1) Is the item stated in a clear and concise manner? and (2) Is a second item buried within the original item? When this situation is encountered it is relatively simple to either disregard the buried item or treat it as a separate item.

Although most of your questions will be straightforward you should be aware of items that apply only to a certain subsample of your sample. For example, you might have a question that requests the graduate to indicate his or her specific academic area of teaching competence. If you want to further investigate his or her particular area of teaching competency through the use of additional items, then subsequent responses will be contingent upon the response to the first item. As you might suspect these types of questions are called "contingency items" and they can be very useful in facilitating the graduate's responses to your survey instrument. The formatting of these items requires creativity and imagination on your part. You must format the items so they do not mislead the graduate while at the same time providing you with an efficient means of coding the responses. The liberal use of arrows and boxes will aid in getting accurate responses.

There are different formats available for allowing the graduates to record their answers on your instrument. You may use brackets, parentheses, boxes, slashes, underlining, and circles. Any of the above methods will get the job done. But one format not to employ is the use of open blanks. You will encounter checks and marks all over the response sheet. This virtually makes it impossible to code the responses.

The ordering of questions is another important consideration in the formatting and coding of your survey instrument. Several rules are in order. Place the "easy" items (demographic data) items at the end of the survey instrument. The more interesting items should be reserved for the beginning. Your intention, as was the spider to the fly, is to entice the graduate to begin responding to the items after a glance at the first few items. He or she should want to complete the instrument.

Although many researchers recommend randomizing the order of questions, we have found it better to order the questions according to some logical scheme. Questions of the same type should be grouped together. One does not think in a chaotic pattern and there is no reason to believe that arranging your items in a chaotic (random) pattern will result in a different response set than if the items were arranged in a logical grouping.

Coding

The proper design of the response format will reduce the probability of transcription or coding errors. Coding is relatively simple for forced responses but difficult or rather cumbersome for open-ended items. The type of coding you use depends on your particular objectives and the

type of processing you are able to do. However, since most schools have a computer or access to a computer, the suggestions are made for the purpose of making your interface with the computer a happy experience. Following is an unrandomized list of suggestions for the coding of your instrument:

1. All of your data items must be compatible with the computer programs and the computer equipment available to you. Before you begin coding, please check with your computer center-- they can be of significant help.
2. Record all responses for each graduate on one card. If this is not possible and you require two or more cards per graduate, you must include the Identification Number and card sequence on each card. This is to prevent a disaster when your deck-of-cards is dropped.
3. Use numbers for codes (1-9). Zero should be reserved for an indicator of "no response." Do not use alphabetic characters.
4. If you are not using mark-sense sheets, you should indicate on the answer forms the card columns containing the variable. If you have an item that requires a yes or no response, then indicate on your survey instrument the value of that variable (1 = yes; 2 = no) and the column where it should be punched.
5. Multiple choice questions with a single response should be arranged for a maximum of nine choices. The choices should be coded 1 through 9 with a 0 indicating no response. For example:

Item 32:

How many hours of education courses do you have? (Col. 41)

- | | | | |
|---------|-------|---------|---------|
| 1 _____ | 0-10 | 6 _____ | 51-60 |
| 2 _____ | 11-20 | 7 _____ | 61-70 |
| 3 _____ | 21-30 | 8 _____ | 71-80 |
| 4 _____ | 31-40 | 9 _____ | Over 81 |
| 5 _____ | 41-50 | | |

A one (1) would be punched for the 1-10 response; a two (2) for the 11-20 response, etc.

6. Another type of multiple choice item is as follows:

Item 37:

Choose three of the following subject areas as the most important to your job? (Col. 42-44)

- | | |
|---------|------------------------|
| 1 _____ | Teaching methods |
| 2 _____ | Educational philosophy |
| 3 _____ | History |
| 4 _____ | Research methods |
| 5 _____ | Learning theory |
| 6 _____ | Audio-visual method |

Item 37 would be represented by three numbers punched in columns 42, 43, and 44.

7. Rank order questions will require one column for each item if you limit your ranks to 9. If you wish more than 9 ranks, then you will have to double the number of needed columns. However, we do not recommend more than 9 ranks. In fact, most ranking questions will have 5 possible choices: such that 1 would represent "the most important" and 5 would represent "the least important."

Item 38:

When choosing an elective course rank the following items in the order of their importance to you when making the decision. (1 = most important; 5 = least important)

_____	Professor	(column 45)
_____	Content of course	(column 46)
_____	Textbook	(column 47)
_____	Relation to your major	(column 48)
_____	Time schedule	(column 49)

8. Questions requiring sorting or grouping for the purposes of data processing should use the digits 1-9. All computer programs cannot sort on the specified variables and the sorting may have to be done by a mechanical card sorter. Each column you want to sort on requires one additional pass through the sorter.

Item 56:

What is your sex?

(Col. 60)

1 _____ Male
2 _____ Female

9. The size of a particular response is limited by most computer programs. The responses of items that require more than a few columns can frequently be transformed into a single number. For example:

Item 58:

What year did you receive your B.A. degree? _____

If the entire number were used, four numbers and consequently four columns would be required. To reduce the year to one

number: remove the first two digits (the century prefix); from the remaining two digits subtract the last two digits of the year the degree was first awarded by the institution and then add 1 to the result, such that the lowest response is one (1). The response to this question may now be used in a frequency distribution.

10. Items requiring a free-response will require a coded list from which you must assign a number.

Item 59:

What is the name of your home state?

(Col. 65-66)

There is one possible response. However, it may be any of the 50 states. You must assign a number to each state and request that the graduate enter that number on his answer sheet or have the graduate write the state on the answer sheet and have your TPAG transform his answer to the proper code. All the codes must be two digits and will require two columns. For example:

STATE CODES

01---Alabama
02---Alaska
03---Arizona
etc.

One last reminder. Avoid using matrix-type items. When the graduate is required to specify numerous conditions as in a matrix-type question, it is often confusing to the graduate and almost impossible to code and subsequently process.

ACTIVITY 2-3**SELECTION OF SAMPLE**

The intention of selecting a sample is to find a group of people who are representative of the population which you wish to describe. In your case the population is the graduates of your institution. When you conduct your survey analysis you do not want to query the entire population because, in many cases, there is a lack of time and the cost would be prohibitive.

Although time and cost are compelling reasons for not using the entire population and given that we are unable to collect population data as meticulously as sample data, it turns out that in many cases the results from the sample are more accurate than the results from the total population. The reasons for this are several in number. First, to conduct a survey analysis of the entire population requires an unusually large staff. Maintaining this large staff is expensive not only in terms of cost but also in terms of accuracy. Processing a large number of returns most likely will result in errors that could possibly bias your results. Secondly, since such a large number of responses would be required to conduct a survey analysis of the entire population, the factor of time would become a variable that might bias the results. Not everyone in the survey would answer at the same time and if the survey took place over a long period of time, this time factor would become critical. Thirdly, the follow-up procedures are more complex. This complexity could result in processing

errors that might bias the returns. It should be noted, as a case in point, that the United States Census Bureau evaluates the accuracy and quality of its population data by conducting a sample survey after the census is completed.

Before we look at different types of sampling designs in this activity, it is important at this point to define a few terms. The reader should be reminded that this is not a definitive discussion of sampling techniques and the ideas presented here are to be used as reminders. For a more theoretical approach and a more accurate approach, one should consult one of the many fine texts available that will guide the reader down the correct path of sampling and selection procedures. Again, we reiterate the point stressed in the introduction--the TPAG should have a statistician on its staff or have one close at hand.

Now, let's look at some terms:

Element: The unit-of-analysis; the basic unit of the survey analysis. The basic unit-of-analysis in this survey analysis is the graduate of your teacher preparation program.

Survey population: The group of elements from which the sample is actually selected. The group of people that we want to describe in terms of their traits and attributes.

Sample unit: The set of elements that are representative of the survey population.

Sampling frame: The actual names of the people from the population from which the sample will be selected. An example would be the list of students registered in the teacher

training program. If you were only interested in describing the population of women students, the sampling frame would be the list of women students enrolled in the teacher preparation program.

Variable: A specific characteristic about the unit-of-analysis that can take on different values or rather has variation. For example, if your population consisted of all the graduates, male and female, that completed your teacher preparation program, then a variable would be sex. However, if your sampling frame was the list of women graduates then sex would not be a variable since all the subjects would be female.

In summary then, you want to derive a sample of elements from a sampling frame that will allow you to describe certain variables of the sample that you can attribute to the survey population.

Types of Sampling Design

There are many different types of sampling designs. For the purpose of deriving an acceptable sample, there are three types of sampling procedures that will serve. The first is called simple random sampling, the second is called systematic random sampling, and the third is stratified random sampling. Notice that each has the word random in it. What do we mean when we say we have a random sample? This means that we have a sample that is representative of the survey population which we want to describe. For a sample to be truly representative, every element of that population must have an equal chance

of being placed into the sample. By stipulating this rather strict rule upon yourselves, you can be assured that your sample is truly representative of the survey population and, furthermore, it allows you to apply some statistical procedures to your data that you would not otherwise be able to.

ACTIVITY 3-5SELECTION OF SAMPLING FRAME

The unit-of-analysis in our evaluation procedure will, in most cases, be those persons graduating from our teacher preparation program. The major problem at this point is to determine the proper sampling frame from which we will draw our sample. Conceptually, this is an easy task. However, in terms of practicality it can be the downfall of the study. This is because accurate lists of graduates are sometimes hard to come by. If your college is maintaining an updated list of graduates and their places of employment, then the task can be accomplished with little difficulty. But if your situation is like most colleges, then locating a proper sampling frame will be a cumbersome activity. In most cases it will be necessary to generate your own sampling frame. We have found, through experience, that various sampling frames available to the college such as registrar's lists, class lists, and alumni lists are too inaccurate for purposes of conducting an effective survey analysis in terms of teacher preparation programs.

One method which has been employed by the University of Arizona appears to have promise in updating and keeping an accurate list of names of education graduates. They have a postcard that is mailed periodically to all graduates listed in their sampling frame. This postcard asks for information such as current addresses and changes

of addresses and position. In addition, they are able to ask a few questions concerning the graduate's job and his satisfaction with the job as a preliminary step to completing the final survey instrument. This use of the postcard has several advantages: (1) it enables you to keep a fairly accurate list of addresses of your graduates; (2) you are able to identify those who have changed positions; (3) you are able to get a general feeling for the graduates' attitudes toward their positions; (4) you are able to tabulate any changes in the employment status of the graduate with regard to staying in education or leaving education; and, finally, (5) it is a relatively inexpensive procedure for maintenance of a proper sampling frame. We highly recommend this procedure.

If the evaluation procedure is to be an on-going procedure which this model assumes, then the TPAG should set up a permanent procedure for generating an in-house list of graduates. This might be done as part of the formal check-out procedures required at some colleges. If your college does not have a formal check-out procedure, you might institute one for the purposes of generating an accurate list. If an in-house procedure is set up, then the maintenance of a comprehensive list of graduates and, consequently, a proper sampling frame is relatively easy and inexpensive.

The maintenance of your sampling frame should be kept on computer tapes, punched cards, or other automated devices. Using the computer allows you to update your list at will and maintain an accurate listing of graduates. It also provides a convenient, accurate method of selecting

your sample. If your sampling frame is maintained on computer tape or other appropriate source, the printing of labels can be accomplished with ease on the computer. For example, if you intend to send out the initial questionnaire, a reminder letter, and two follow-up questionnaires, then you can have the computer print four sets of labels for each person selected to be in the sample. If you do not have anyone in your TPAG who is familiar with computer activities, we suggest that you include one or, like the statistician, have one close at hand.

In summary, then, we recommend the following:

1. Set up an in-house procedure for compiling an accurate list of graduates. The list of graduates must be compiled at the point of graduation and subsequently updated at each graduation point.
2. The list must be periodically updated. The use of a simple follow-up postcard system is recommended.
3. Maintain your sampling frame on the computer (if possible).
4. Maintain a close contact with your computer center personnel.
5. Have computer programs written to update your sampling frame and to execute the printing of a prespecified number of labels.

This activity is concerned with the choice of a proper sampling procedure to insure that a representative sample is chosen for the survey analysis. Three methods will be discussed: (1) simple random sampling, (2) systematic sampling, and (3) stratified sampling.

Simple Random Sampling

Simple random sampling, as the name implies, is the simplest method of selecting a sample from your sampling frame. The general rule to keep in mind about random sampling procedures is that every individual of the population has an equal chance of being selected as part of the sample. Also the selection of any one individual of the population must in no way relate to the selection of another. Random sampling is necessary to assure a representative sample of the population. Also, the use of sampling statistics such as standard error, etc. depends upon the assumption that sampling has been a random procedure.

To select a sample using a simple random sampling, it is only necessary to assign consecutive numbers to each member of your sampling frame and then, through the use of a random number table, select the sample. If your sampling frame is in the computer or in computer readable form, then the computer can be programmed to not only number the members of the sampling frame but to select them in a random fashion. This is a relatively easy task for the computer.

Systematic Random Sampling

Systematic random sampling is a method that is used more frequently than simple random sampling because, if done by hand, simple random sampling is an arduous task. In addition, every member of the sampling frame must be available for selection. Systematic random sampling involves the selection of every n th element in the sampling frame for inclusion in the sample. It is possible to start with the first element and then choose every n th element until the sampling frame has been exhausted. However, this introduces human bias into the procedure and could destroy the representativeness of your sample. The proper procedure is to choose the first element at random and then select every n th element thereafter. This is called systematic random sampling with a random start.

If your sampling frame consisted of 1000 names and you wanted a sample of 100 subjects, you would choose the first number from between 1 and 10 at random and then select every 10th (n th) element for inclusion in the sample. Systematic random sampling is simple and easy to do. All you have to do is to select the sampling interval (in the above example this is 10) and the sampling ratio which is the ratio of total elements in the sampling frame to the number selected for the sample. The sampling ratio in the example is 10:1.

Stratified Random Sampling

Stratified random sampling is a modification of simple random sampling and systematic random sampling procedures. To obtain a stratified sample from your sampling frame, you must first organize your

elements into groups having the same characteristics, traits or attributes that are homogeneous in nature such as academic disciplines, departments, classes, sex or any other homogeneous grouping that has meaning for your particular survey analysis. The procedures for selecting a sample from each of these subgroups is the same as simple random sampling or systematic random sampling. The important thing to remember is that the subgroups are homogeneous within themselves, but heterogeneous between the subgroups. Stratified random sampling is a desirable method because it is amenable to computer processing and has the added advantage of producing a smaller sampling error than the other methods when the stratification procedure is used on all relevant independent variables.

ACTIVITY 9-12

DETERMINING SAMPLE SIZE

Many people at this point ask the question, "How large should my sample size be?" The answer to this question is easy--Make it as large as possible.

As a general rule one can expect more precision in estimating the population parameters when a larger sample is used. The diagram in Figure 2 gives a general idea:

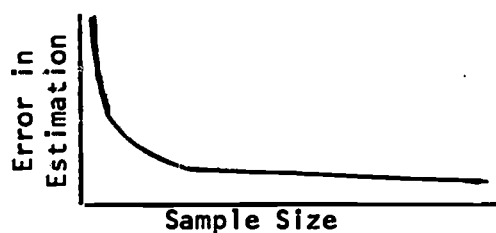


Figure 2.--Relationship of Sample Size to Error.

As the sample size increases, the error decreases and, therefore, the precision in estimation increases. One must be cautioned not to carry this to the extreme of including the entire population in the sample for if this were to happen there would be no need to sample and, hence, no reason for the use of sample statistics. It can be shown that after a certain sample size is reached, a large increase in that sample size does not result in a compensating decrease in error. From a practical standpoint (cost), the use of extremely large samples will not offer the promise of more accurate results.

Nevertheless, under certain conditions, large samples are necessary. These conditions exist when (Isaac & Michael, 1971):

1. there is a possibility of a large number of uncontrolled variables interacting in an unpredictable manner;
2. subgroups are the units-of-comparison and the total sample must be divided into these subgroups;
3. the population is comprised of a wide range of characteristics and variables; and
4. differences in the results are expected to be small.

The real question to be answered is not how large a sample to choose, but how small should the sample be to insure adequate representation of the population?

The answer is again easy--one. Only one member of the population is needed to describe the entire population. You may scoff at a sample size of one, but consider how many pieces of lead you would need to describe the properties of lead. You would most certainly not require all the lead in the universe to give an adequate description of its attributes--just one small piece. Maybe only one molecule. The problem that faces us in educational research is that, unlike lead, each person is different and the examining of just one graduate would bias our conclusions if that one graduate was not truly representative of our survey population of graduates. In educational research we are concerned with the differences or variations in our graduates. We should not try, as is frequently the case, to decrease this variation, but to design a teacher preparation program to take advantage of these variations.

For, as William Cowper (1731-1800) said, "Variety's the very spice of life." Our education system should not try to deny this.

Since the choice of just one person will obviously not allow us to adequately describe our survey population, we must choose a sample of an appropriate size from our population. We will not delve into the technical and theoretical constraints involved in choosing an appropriate sample, but we will offer some guidelines and present an equation that might be useful to you in determining the proper sample size for your particular situation.

Confidence Intervals and Levels of Confidence

As we have said before, the immediate issue is to choose a sample size that insures an adequate representation of the population. Although it is never possible to be completely sure of the adequacy of your sample, you can be sure within certain confidence intervals.

The point in specifying a confidence interval is to remember that as the confidence interval increases, the larger sample size you will need to assure representativeness. Levels of confidence are usually set at the 95 percent level which means that if you select 100 samples from your sampling frame, the characteristics of 95 of the samples will be distributed as your population characteristics. Put another way, five of your samples will not be truly representative of your population. Other levels may be specified such as the 99 percent level, the 90 percent level, or any level of confidence you are willing to accept.

Sampling Error

Another consideration is sampling error. Sampling error refers to the degree of error that exists between your sample and the population characteristics. If the population-mean of a certain trait is equal to u_i and the sample-mean of that trait is X_i , then the sampling error is the difference between the population-mean and the sample-mean: such that $e_i = u_i - X_i$ where e_i is the sampling error for a particular trait. In determining sample size then, we are interested in keeping the sampling error as low as possible.

An equation for determining how small a sample should be is given in a 1960 National Education Association Research Bulletin (NEA Research Bulletin, 1960). It is as follows:

$$n = [\chi^2 N \pi (1-\pi)] / [d^2 (N-1) + \chi^2 \pi (1-\pi)] \quad \text{EQ. 1}$$

where n = the required sample size

χ^2 = the table value of chi-square for one degree of freedom and desired confidence level. (For 95% confidence level $\chi^2 = 3.841$)

N = the population size

π = the population proportion which it is desired to estimate

d = the degree of accuracy expressed as a proportion

Example: Consider the following: Your secondary education graduates accounted for 50 percent of all your graduates. How many secondary education graduates would you sample if you were willing to accept a confidence level of 95 percent and a sampling error of .03 for your population of 500 graduates?

Equation 1 becomes:

$$n = [3.841 \cdot 500 \cdot (.5)(1-.5)] / [(.03)^2(500-1) + 3.841(.5)(1-.5)]$$

where $\chi^2 = 3.841$

$$N = 500$$

$$\pi = .5$$

$$d = .03$$

$$\begin{aligned} \text{then } n &= \frac{480.13}{.4491 + .9603} \\ &= \frac{480.13}{1.4094} \\ &= 340 \end{aligned}$$

Other formulas are given in many texts for estimating the proper sample size (Guilford, 1965; Tuckman, 1972). For those who do not want to use the equation, tables are available. A table of sample sizes required for finite populations is given in Tables for Statisticians (Arkin & Colton, 1963). Another set of tables is available from the National Education Association Research Division (National Education Association, 1965). These and other available sources will help you in estimating the proper sample size for your particular situation.

ACTIVITY 7-10**DEVELOP PROTOTYPE INSTRUMENT**

When the prototype instrument is developed you must be concerned with the instrument in terms of internal consistency, the cosmetics of the instrument, and the logistics of administering it.

Internal consistency refers to the logical agreement between instructions, instrument format, response modes and relation of items to objectives. You should work through the instrument as though you were the graduate. You can be sure that if you encounter problems in completing the questionnaire, then the graduates will. This check of internal consistency goes far beyond a mere visual inspection of items. You must insure that the items are related to the objectives, remove those items that are not, and add items that are necessary. Caution must be exercised when any of the items are changed to insure that in the process of changing and reformating, needed information is not deleted.

Cosmetics refers to the general appearance of the questionnaire. Wide margins and liberal spacing between items should be used throughout. If items are packed together you increase the probability of confusion. This is critical not only for closed items but also for open-ended items. When an open-ended item is presented with a lack of space to answer, you can expect incomplete responses, illegible responses, or responses that are continued to other parts of the instrument. Certainly any of these

occurrences will reduce the accuracy of your responses and, consequently, your overall data base.

A final point relates to the professional appearance of your instrument. Do not sacrifice quality for the savings of a few dollars. We have seen many survey instruments that are respectable in all matters except for the general appearance. Questions are squeezed together, the type is too small and, in many cases, difficult to read. This variable in itself plays an important part in the number of returns and the quality of the responses.

Reproducing the prototype instrument on ditto or mimeograph is acceptable for the initial tryout procedures, but we do not recommend these types of reproductions for the final instrument. Although they are cheap and readily available, they do not provide a quality product with a professional appearance. The use of offset lithography procedures is recommended for the final product. It is not only professional looking, but also it can even be cheaper if you have a large sample size.

There is one other device that we have found to be invaluable in producing our survey instruments. This device is the IBM Magnetic Card (Mag-card) typewriter or similar device. These machines record your typing on magnetic cards or magnetic tape. This allows for easy editing and reformatting since you can delete or insert characters, words, sentences, or larger portions with ease. The machines will then play back (type) at high speeds mistake-free finished copy. These machines are well worth the investment.

In a prior activity the use of a common denominator for grouping was discussed. You should review your prototype instrument to insure a logical grouping of items. Do the instructions match the appropriate groupings and are they clear? Try your prototype on a few people. Can they successfully complete it without your help? Where do they need help? What are the problems in administering the instrument? Now is the chance to answer and correct these problems. Do not wait until the instrument has been returned from the printer in its final format. Do it now!

In summary -- REVISE AND REFORMAT, REVISE AND REFORMAT, REVISE AND REFORMAT. When you have revised and reformatted your instrument to correct all the observed problems, REVISE AND REFORMAT again.

ACTIVITY 10-11**PILOT TEST AND REVISION**

Pilot testing is a means to "debug" your survey analysis instrument. Pilot tests are usually run on a small sample of graduates selected from your sampling frame but who will not be used in the final sample. Although it is nice to have a "representative sample," sophisticated sampling techniques are not necessary. You are not interested in obtaining "statistically significant" results, but in obtaining information relating to the validity of your instrument in terms of your objectives. In other words, do the items measure what you want them to measure?

The responses your pilot group give can be extremely useful in determining deficiencies in your instrument. You should examine the results of your pilot run with an eye toward looking for items that have the same responses from all or most all of your pilot sample. When you discover these type of items, you should either discard them or reword them because they do not discriminate. It is useless to have an item that everyone will give the same response to (this does not include items that request biographical or demographical information). If you have items that elicit inappropriate responses, examine the items for poorly worded or ambiguous questions or statements.

Some questions may elicit no responses. These items should be examined to determine if they are dealing with sensitive social,

political, or personal behaviors. If they are important to your study, you may be able to reword them so that they will not be offensive to members of your sample.

Statistical techniques can be used to determine if your items are really measuring what you intend to measure. If you have a series of items that are intended to assess the same variable, it is possible to correlate the responses to these items with the mean score across all similar items. The greater the correlation between the item response and the total response, the greater the relationship between what the item is measuring and what the total list of items are measuring. Those items that have a high correlation would be retained and those that have a low correlation would be discarded. You would then have a set of items that you could be assured were measuring the same attribute.

This tryout and revision phase should be conducted until you are satisfied with the results. Are the instructions clear? Poor instructions will become apparent and should be appropriately changed. Are the items unambiguous and, most important, do the items elicit responses that are relevant to your objectives?

ACTIVITY 11-12**DEVELOPMENT OF FINAL INSTRUMENT**

You are now ready for the preparation and reproduction of the final instrument. This includes development and specifications of the survey analysis instrument, envelopes, cover letters, answer sheets, and all accompanying material.

The first area of concern is the survey instrument. In Activity 7-10 we discussed the advantages of an instrument which has a professional appearance. Although cost considerations will vary from institution to institution, we cannot overemphasize the fact that initial instrument impact is a variable which does have an influence on return rate. You should not allow cost to dictate the production of a shoddy instrument. Saving a few dollars here may jeopardize your entire effort.

If you are dependent upon the action of outside agencies, i.e. printing companies, you must allow sufficient slack time in your activity sequence to compensate for possible delays attributable to these agencies. On the other hand, you should be aware of the possibility of internal delays within your own group. Either of these sources of delay have the potential for halting your project.

In contracting with an outside printing agency, you should request a pre-copy run of your instrument. Once this copy has been reviewed by your group, a completion commitment should be obtained from the outside agency. Depending on the project deadline, a job initiation date may also be requested from the outside agency.

The request for a project initiation date and a completion date is essential if the TPAG is operating under a tight deadline for initial mailout. First, your TPAG can contact the outside agencies on those prespecified dates to confirm the initiation of the job. This is a hedge to reveal problem areas before completion dates have passed. Also if time is of the essence, a partial shipment may be obtained to enable TPAG personnel to initiate assembly of survey packages. This again may prevent a bottleneck.

A concurrent activity is the preparation of the outgoing and return envelopes. This activity is composed of three phases. The first phase involves a dry run using materials which have the essential characteristics of the actual materials. For example, a sample survey package should be assembled. This reveals the compatibility of materials to envelopes, etc. It also provides an indicator of the amount of time which will be required to assemble the materials. The second phase is the preparation of envelopes. This involves the printing of postpaid return envelopes and outgoing envelopes. Prior to the delivery of prototype envelopes, the post office should be consulted. This is necessary due to very specific requirements of postpaid mailings. Once

the envelopes have been checked and found acceptable, the sample package and cost estimates can be determined. The final phase is the delivery of the sample envelopes to the printer. It is recommended that a sample set be obtained from the printer and again submitted to the post office for inspection. After they have approved it, the printing of the envelopes should be initiated. Again, a commitment for completion and project initiation date should be obtained from the outside agency.

Before all materials have been put into final form, it is advisable to determine and try out the procedure for physically getting everything together. It can be very embarrassing after everything has been put in final form and produced in quantity to discover that the envelopes are too small or that a cover letter is missing. Your group should go through the actual procedure of collecting and collating all the materials; and, then, act as the graduates by opening the package, actually completing the instrument, and returning it according to instructions. Although this task may seem trivial, it is extremely important. Many researchers, to their regret, assume that everything will automatically go together. Don't count on it--Plan for it!

ACTIVITY 4-8

SELECTING STATISTICAL ANALYSIS AND PROCESSING OF DATA

Needless to say, the processing of the data, the analyzing of the results, and the interpretation of these results is of utmost importance to the completion of your survey analysis. There are so many statistical procedures available for the processing of your data that it is beyond the scope of this discussion to provide a detailed analysis of the various procedures.

Your analysis procedures can conceptually be broken down into two phases: the descriptive phase and the inferential phase.

Descriptive Phase

The descriptive phase is concerned with a presentation of the data without any reference as to why the results were obtained. The purpose of the descriptive phase is to summarize and condense the data into a more manageable form. You want to describe your data in terms that will allow a presentation of it in an understandable fashion.

Some of the descriptive statistical methods are:

A. Calculation of measures of central tendency

1. Mean
2. Median
3. Mode

Each of these measures has its advantages and disadvantages but all will provide an accurate measure of the "average" in the appropriate situation.

B. Calculation of measures of variability

1. Average deviation
2. Absolute deviation
3. Standard deviation
4. Variance
5. Range

These measures give some indication of the "spread" or variation of the responses to the items. The most common measure of "spread" is the standard deviation. The purpose of measures of variability is to give the researcher some idea of the "spread" of his scores about some central point. To specify just the mean or "average" score has little meaning unless you combine it with a description of the variation of the scores about that "average." Therefore, you should always report your average with a description of the variability of the scores about that average.

The primary function of the descriptive phase analysis is to present your data in a meaningful manner. If you were to report the responses of each graduate on each item of your survey instrument, the presentation of the data would be unmanageable. Subsequent interpretation would be almost impossible. The use of descriptive statistics provides an efficient facility for reducing large amounts of data to meaningful summaries. You may summarize your data in terms of percentages, frequency distributions, or logical groupings of your data into some meaningful manner with the subsequent use of the previously discussed measures of central tendency and variability. The key point to remember in using descriptive techniques is that you have a trade-off point in the summarization process. Although we advocate as much summarization as possible, we do not recommend summarizing your data to

the point where the real meaning of the data is lost. Your goal in the descriptive phase is efficiency in terms of maintaining the maximum amount of information with the minimum number of data points.

Inferential Phase

The inferential phase is concerned with inferring traits and attributes about your survey population from the measured traits and attributes of your survey sample. In the descriptive phase you were interested in defining whether a relationship existed between two or more variables. In the inferential phase you are interested in describing why the relationship or non-relationship exists.

There are many statistical procedures for you to use in making inferential statements about your particular population. Some of these are: Chi-Square Analysis, t-Test, Analysis of Variance, Analysis of Covariance, Multiple Regression Analysis, and Factor Analysis. All of these procedures have been used quite extensively in educational research and there are many references to the use of these procedures.

One statistical analysis procedure that has not seen much use in educational research, up to this point, but is becoming increasingly more popular as time goes on is the use of Bayesian Statistical Procedures. If you are one of those researchers that likes to try new methods, then you might be interested in investigating the applicability of using Bayesian Statistics to analyze your data. An excellent reference is Statistical Methods for Educational and Psychological Research (Novick & Jackson, 1974). Novick and Jackson furnish other references that will

provide additional information regarding the use of Bayesian Statistical procedures.

Getting a Feeling for Your Data

Although we have emphasized the use of standard statistical procedures, there is one other procedure that is not written about in the textbooks. This procedure is what we call "sleeping with your data." The general idea, as you might guess, is to gain an intimate knowledge of your data. Having an intimate relationship with your data can provide you with a wealth of information that is otherwise unavailable through the use of standard statistical procedures. Although you cannot stipulate that the relationship between variable X and variable Y is statistically significant at the .05 level of confidence, you can say there may be a relationship between the variables and further investigation should be conducted. This getting-to-know-your-data procedure can provide insights into possible problems and solutions that would never become apparent through the use of standard statistical analysis. We are not saying do not use the standard statistical procedures, but don't rely on them entirely. Have an affair with your data!

Computer Processing

You will no doubt use the computer for the initial processing of your data and subsequent analysis. Before you say "the computer will do it" make sure your computer has the capability to perform the functions you want. You should evaluate the available resources and make arrangement with your computer center ahead of time to process your data.

There are many commercially prepared programs available to process your data, but you must determine in advance if your computer center has them available. One computer package that is extremely useful is called the Statistical Package for the Social Sciences (SPSS) (Hull, 1970). The proper use of the SPSS computer package can save you countless hours of work. If it is available--use it. Another package is called the Biomedical Computer Programs (BMD) (Dixon, 1971). The BMD package is not as versatile as the SPSS package, but it is easy to use and can be very useful. Both of these packages are listed in the reference section.

If these packages are not available to you or you are inclined to write your own programs, you might want to consult Fortran Programming for the Behavioral Sciences (Veldman, 1967). This book includes an introduction to the computer language FORTRAN with many statistical routines. It is an excellent source on programming statistical routines in FORTRAN.

Although computers are necessary to process your data and the computer programs that are available to you will, in most cases, give you more information than you will need, you should not blindly accept the results handed to you by the computer. Always spot check your results by hand. We have found too many computer programs to be in error and have seen too many researchers or so-called researchers accept the results on a computer printout as if it were a message from God. A statement that is hanging in our office is a reminder of the monumental problems that an incorrectly programmed computer can cause. It is:

"It would take 100 clerks working 100 years
to make a mistake as monumental as a computer
can make in 100th of a second"¹

The computer is a wonderful machine, but it is only as good as
the people who program it and use it intelligently.

¹ Author unknown

ACTIVITY 8-12**PROCEDURES FOR PROCESSING RETURNS**

The purpose of this activity is to set up procedures to handle the survey instruments as they are returned. Two objectives should be kept in mind for the successful completion of this activity: (1) You want to eliminate those graduates from the survey sample who have completed the instrument so they do not receive follow-up mailings; and (2) The returned instruments should be sorted into logical groupings for efficient data processing.

Check-Off Procedures

This section deals with the materials and procedures for keeping track of who responded. The check-off procedure has two prerequisites: a master list of the graduates who were included in the mailing and a means of identifying the graduates who have returned the questionnaire. Typically, the desired number of respondents is not achieved within one mailing. Consequently, follow-up mailings are necessary to the non-respondents. The master mailing list must contain the name and address of all the sample graduates. In addition, there must be an identification section to match returned questionnaires to graduates on the mailing list.

There are two approaches for implementing an identification section: (1) use available or existing coding systems and/or (2) deriving a code specifically for the project. There are a number of codes which

already exist for identifying individual graduates. The most obvious is the graduate's name, but it should be noted there are a number of serious disadvantages to this approach. Most important, the success of the system depends entirely on the graduate including his name on the returned questionnaire. In other words, if the graduate omits this information, it is impossible to determine who responded. The use of an alphabetic code also presents problems in the areas of transcription and analysis. Generally, alphabetic information is either impossible or difficult to process when the computer is used to recode the information. In addition, statistical packages generally are not designed to manipulate alphabetic information. Consequently, the use of the graduate's name is not an efficient method for identification.

There are also a number of existing numeric codes which are unique for each student (i.e., social security number, matriculation number, etc.). The procedure involves coding each questionnaire with the appropriate social security number and then when the questionnaires are returned matching the social security number with the corresponding one on the master sheet. This is very cumbersome and, because the code was developed for another purpose, it is very costly for the benefit received. In other words, the social security number only identifies a unique individual and requires most of the work of a code contrived specifically for the project without the additional information provided by a contrived code.

A contrived code is a means of identifying and classifying individuals into meaningful categories. This code may be as simple as a

serial number with a prefix for department membership (or other classification variable). Its advantages lie in the efficiency of the check-off procedures. There is no reliance on the graduate to input the information and the code can provide more than the identification of the graduate.

The check-off procedure involves identifying the graduates who completed the questionnaire and those who could not be reached (assuming the instruments were mailed with first class postage). In other words, when eliminating names from the follow-up list, two codes are necessary: one for those completed by the graduates and one for those returned by the post office due to an inaccurate address.

Once the questionnaires have been logged into the master list, the next step is to sort them into categories which will facilitate punching or coding. For example, portions of the identification section will remain constant for a given department and by sorting them on this variable the encoding is streamlined.



ACTIVITY 12-13**INITIAL MAILING**

The distribution of materials is contingent on the completion of three tasks which have been occurring simultaneously: (1) the return procedures (Activity 8-12) which specifies tasks required for sorting and cross out procedures; (2) production of final instrument (Activity 11-12) which includes assemblage of survey packages; and, finally, (3) the determination of graduates and sample size (Activity 9-12).

Generally, this is a straightforward procedure; however, the post office may require certain packing techniques. If a postage meter is not available you might consider obtaining one. Each type of package must be weighed to insure proper postage. Don't allow your survey analysis to fail because of improper postage.

First class mailing is recommended for outgoing instruments for three reasons. First, this insures that if an individual moves the instrument will be forwarded. This is especially important since many of the graduates of teacher education programs are young and mobile. Secondly, when an addressee cannot be located the instrument is returned to the sender. This is important for determining response rate. Also, it provides a feedback loop so that follow-up mailings are not sent to unattainable individuals. Finally, it provides a list of individuals who are not part of the current survey population. These names should be eliminated from the sampling frame.

ACTIVITY 13-14

KEEPING TRACK OF RETURNS--CHARTING AND FOLLOW-UP

The questionnaires have been mailed and there is nothing to do until all the questionnaires have been returned. Right? Wrong! The TPAG should undertake the procedures listed in Activity 8-12 regarding the set up procedures for handling the returns. If Activity 8-12 is successfully implemented, then the charting procedures outlined here are relatively easy.

If one is to meet with success in terms of getting the most out of the questionnaires, keeping track of the returns is extremely important for making decisions regarding when to mail the second follow-up and when to cut off acceptance of returns. Knowing when to mail follow-ups and knowing when to terminate acceptance of returns can be extremely worthwhile in terms of man-hours and, eventually, cost. It is expensive to keep a crew on board waiting for returns that will never come.

Although we do not have a magic formula for determining when to follow-up and when to cut-off, there are some procedures and general guidelines that will assist you in making the proper decisions. A tool that is a must is frequently referred to as a return-rate-graph. As the name implies, this type of graph keeps track of the rate of returns. Two types of graphs should be prepared: (1) a graph to indicate the number of daily returns, and (2) a graph to indicate the percentage of returns. Examination of these two graphs will provide the TPAG with

information as to when to initiate a follow-up mailing and when to cut off acceptance of the returns.

Graph No. 1, showing the number of daily returns, includes the days labelled along the X-axis and the number of returns labelled along the Y-axis. Day No. 1 is the day the questionnaires were mailed. Then, on every subsequent day the number of returns is recorded on the graph.

Graph No. 2, showing the percentage of daily returns, provides a cumulative record of the number of returns. Again, the days are labelled along the X-axis but the percentage of returns is labelled along the Y-axis. As the questionnaires are returned, the number of returns in relation to the total number mailed or the percentage of returns is recorded on the graph for each day.

By following the ups and downs on the first chart and the rising percentages on the second, one is able to determine when the returns begin to slacken. If the first chart shows a definite decline in the rate of responses and the second chart shows a constant rate of percentage of returns, it is then time to initiate the follow-up and mail another questionnaire. Some researchers feel that a follow-up letter is sufficient, but we would recommend that another questionnaire be mailed. The follow-up letter is useless if the graduate has misplaced the original questionnaire and, as you well know, this is extremely easy to do. In addition, being faced with the questionnaire may just spur him to complete it rather than filing the letter and, consequently, your questionnaire. As a rule of thumb, the longer a graduate delays

in responding, the less the probability is that he will return the original questionnaire.

Although there is no statistical rule to follow in determining when to initiate the follow-up mailings, Babbie (1973) offers some good advice:¹

1. Three mailings (an original and two follow-ups) is most efficient.
2. Within two weeks after the first mailing 40 percent of the returns should be returned.
3. Within two weeks after the mailing of the first follow-up, an additional 20 percent should be received.
4. Within two weeks after the final follow-up mailing, an additional 10 percent should be received.

It should be noted that Babbie (1973) cautions the reader not to assume that a similar pattern could be expected in surveys of different populations. However, the guidelines do serve to emphasize the value of carefully recording the number of daily responses and the daily response rate for each survey.

Response Rate

We attempted to determine just what an acceptable response rate is and found as many different answers as people we asked and the publications we read. However, there appears to be a consensus of sorts

¹Determined by the Survey Research Office, University of Hawaii, from student surveys.

and, again, Babble (1973) provides us with a livable guideline. He says:

a response rate of at least 50 percent is adequate for analysis and reporting. A response rate of at least 60 percent is good. And a response rate of 70 percent is very good. The reader should bear in mind, however, that these are only rough guides, they have no statistical basis, and a demonstrated lack of response bias is far more important than a high response rate.

Formula for Response Rate

To calculate response rate the following formula is given:

$$RR = [q/(N-U)] \times 100 \quad \text{EQ. 2}$$

Where: RR = Response rate
 q = Number of returned questionnaires
 N = Number of initial questionnaires mailed
 U = Number of undeliverable questionnaires

If we had an initial sample size of 200 with 30 of these returned as undeliverable and 120 completed questionnaires returned, we would have a response rate of 70.59 percent.

With: N = 200
 U = 30
 q = 120

Then: $RR = [120/(200-30)] \times 100$
 $= [120/170] \times 100$
 $= .7059 \times 100$
 $= 70.59\%$

This formula is based on the assumption that the questionnaires that were undelivered were a random sample of the set of initial questionnaires.

This activity is concerned with the methodology used in the summarization and analyzation of the responses. There are three approaches to achieving this goal: manual, manual-automated, automated. The decision of how to accomplish this is based on sample size, type of analysis to be performed, and available resources.

The manual approach is feasible only if the sample size is small and the analyses are relatively simple. This approach requires constant human intervention and manipulation of the data. This has two primary disadvantages: first, the dedication of personnel to the tedious task of number crunching; second, the inaccuracies of data manipulation due to the constant human intervention. Consequently, you should evaluate the costs encountered by this approach both in terms of expended man-hours and accuracy.

The manual-automated approach involves the integration of manual operations and computer operations. Typically this involves the recoding of the source document into a computer compatible format. This may be in the form of keypunching and verifying the data before input to the computer for analysis. This not only increases the accuracy of the analysis, but also enables a time reduction in the analysis of the data. The principal disadvantages of this approach are costs, the need for skilled personnel, and loss of accuracy due to human intervention.

The costs of the use of a computer may not outweigh the advantages of speed and increased accuracy when the sample size is small and the analyses are simple. The requisite of skilled personnel to keypunch, verify, and process the data is also an added cost. The third disadvantage involves the loss of accuracy via the transcription and punching process. Although this approach reduces the amount of human intervention, the problem is still present but to a lesser degree.

The fully automated approach typically involves the use of an instrument keyed to a mark-sense answer sheet. This process offers the highest degree of accuracy with the lowest error rate. In addition, it is the fastest in terms of time required for analysis. The disadvantages are twofold: limitation of question format and costs. The limitation on question format deals primarily with the fact that all questions must be stated in a forced response mode. Secondly, the costs of using the mark sense answer sheets involve such things as mailing envelopes which prohibit folding of the sheets, printing special sheets if anything other than a standard format is used, and the costs incurred in using the reader (mark sense, optical scanner), computer system, and skilled personnel to operate them.

Again, the objectives of your study, the availability of human resources, the availability of computer resources, and the amount of money you have will determine which procedure or combination of procedures you will use. The proper mix of manual with automated methodology is a function of the available resources, but it should be noted that

inaccurate results are a direct function of the magnitude of human intervention. Due consideration must be given to the cost vs. accuracy tradeoff.

PREPARATION OF FINAL REPORT

The title of this activity is self explanatory. The primary purpose of this report is to provide feedback to the curriculum developers in your college regarding the quality and effectiveness of your teacher preparation programs. An additional purpose is to provide your Teacher Preparation Assessment Group with data that will allow them to improve their measuring instruments and assessment techniques. To accomplish these tasks the report must be written with regard to the intended audience. Data should be summarized and presented in a manner that is at the same time both complete and concise.

In presenting your data you should give enough information so the reader can recompute important statistics. It is a good report that will provide the reader with enough information to enable him or her to replicate the study if he or she so chooses.

Do not relegate your tables, etc. to an appendix of your report. This requires the reader to turn back and forth between the text and the appendices trying, in many cases, with much frustration to locate the proper table. It may be easier for you and the typist, but placing the tables in the appendix is a roadblock to the reader. The tables, graphs, or charts should be placed as close to the referent text as possible. You should first introduce the table and the purpose for

presenting it; insert the table; and, finally, provide a concise review and interpretation of it.

Since your study has both a descriptive purpose and an explanatory purpose, you must provide the reader with enough information to allow him a solid basis for determining relationships between the variables that are of interest to him. Of course, you will provide explanations of most of the relationships, but the presented data should be sufficient for further exploration by the reader if he or she so chooses. In addition to presenting descriptive data and explanations regarding relationships, you want to provide the reader with information that will allow him or her to propose a course of action. Again, your final report will provide suggestions and recommendations, but your data must be in such a form that the reader can make additional recommendations after careful consideration of the presented data.

Key your discussion of the results and conclusions to your originally stated objectives. Did your study, in fact, answer the questions put forth by your objectives? Do the conclusions support your initial hypotheses? In other words, did your study accomplish what it was intended to do?

REFERENCES

- Abramson, D. A. Curriculum research and evaluation. Review of Educational Research, 1966, 36, 388-395.
- Academy for Educational Development. A guide to professional development opportunities for college and university administrators. Compiled and edited by J. T. Irwin. Washington, D. C.: Author, 1974.
- Aiken, W. M. The story of the eight-year study. Adventure in American Education, Vol. I. New York: Harper & Brothers, 1942.
- Alciatore, R. T., & Eckert, R. E. Minnesota Ph.D.s evaluate their training: A study of the relationship of various Ph.D. programs to later career service and satisfaction. University of Minnesota, 1968. ERIC #ED 053 695.
- Amidon, E. J., & Hunter, E. Improving teaching: Analyzing verbal interaction in the classrooms. New York: Holt, Rinehart & Winston, 1966.
- The Arizona Republic. May 26, 1974, p. B11.
- Arkin, H., & Colton, R. R. Tables for statisticians. New York: Barnes & Noble, 1963.
- Arlt, G. O. New trends in graduate study in the humanities. In E. Walters (Ed.), Graduate education today. Washington, D. C.: American Council on Education, 1965.
- Babbie, E. R. Survey research methods. Belmont, CA: Wadsworth, 1970.
- Bigelow, K. W. Foreword. In E. V. Hollis, Toward improving Ph.D. programs. Washington, D. C.: American Council on Education, 1945.
- Cartter, A. M. An assessment of quality in graduate education. Washington, D. C.: American Council on Education, 1966.
- Chamberlin, D., Chamberlin, E. S., Drought, N. E., & Scott, W. D. Did they succeed in college? The follow-up study of the graduates of thirty schools. Adventure in American Education, Vol. IV. New York: Harper & Brothers, 1942.
- The Chronicle of Higher Education, Vol. VIII, May 28, 1974, pp. 1-2.
- Conant, J. B. Introduction. General education in a free society: Report of the Harvard committee. Cambridge: Harvard University Press, 1958.

- Cooperation of general education. A Final Report of the Executive Committee of the Cooperative Study in General Education. Washington, D. C.: American Council on Education, 1947.
- Cowley, W. H. 1963: Critical decisions. In G. K. Smith (Ed.), Twenty-five years 1945-1970. San Francisco: Jossey-Bass, 1970.
- Deming, W. E. Sample design in business research. New York: Wiley, 1960.
- Dewey, J. Educational values. Democracy and education: An introduction to the philosophy of education. New York: Macmillan, 1916.
- Dixon, W. J. (Ed.). BMD (Biomedical Computer Programs). Berkeley: University of California, 1971.
- Dressel, P. L. Evaluation in general education. Dubuque, Iowa: Wm. C. Brown, 1954.
- Dressel, P. L. Evaluation of the environment, the process, and the results of higher education. In A. S. Knowles (Ed.), Handbook of college and university administration. New York: McGraw-Hill, 1970.
- Dressel, P. L., & Mayhew, L. B. General education: Explorations in evaluation: The final report of the cooperative study of evaluation in general education of the American Council on Education. Washington, D. C.: American Council on Education, 1954.
- Dressel, P. L., et al. Institutional research in the university: A handbook. San Francisco: Jossey-Bass, 1971.
- Durant, W. The story of civilization: Part I: Our oriental heritage. New York: Simon & Schuster, 1954.
- Dyer, H. S. School evaluation: A realistic response to accountability. The North Central Association Quarterly, 1972, 46.
- Eble, K. E. The recognition and evaluation of teaching. Salt Lake City: Project to Improve College Teaching, The American Association of University Professors, & the Association of American Colleges, 1970.
- Elam, S. Performance based teacher education: What is the state of the art? American Association of Colleges for Teacher Education, 1971. ERIC #ED 058 166.
- Ellis, E. Preface. In F. C. Abbott (Ed.), Faculty-administration relationships. Report of a Work Conference, May 7-9, 1957, sponsored by the Commission on Instruction and Evaluation of the American Council on Education. Washington, D. C.: American Council on Education, 1958.

- Forehand, G. A. An evaluation system for curriculum innovation. Teachers College Record, 1971, 72, 577-591.
- Giles, H. H., McCutchen, S. P., & Zechiel, A. N. Exploring the curriculum: The work of the thirty schools from the viewpoint of curriculum consultants. Adventure in American Education, Vol. 11. New York: Harper & Brothers, 1942.
- Guilford, J. P. Fundamental statistics in psychology and education. 4th ed. New York: McGraw-Hill, 1965.
- Guilford, J. P. Psychometric methods. New York: McGraw-Hill, 1954.
- Guttman, L. A basis for scaling qualitative data. American Sociological Review, 1944, 9, 139-50.
- Hartnett, R. T. Non-traditional study: An overview. In S. B. Gould & K. P. Cross (Eds.), Explorations in non-traditional study. San Francisco: Jossey-Bass, 1972.
- Harste, J. C. Toward an approach to program evaluation in teacher education. Indiana University, Bloomington, 1972. ERIC #ED 060 023.
- Hawkrige, D. G. A summary of the instructional system used by the Open University in Great Britain in 1971. The Open University, 1971. ERIC #ED 060 605.
- Heath, R. W. Curriculum evaluation. In R. L. Ebel (Ed.), Encyclopedia of educational research. London: Macmillan, 1969.
- Heiss, A. M. Challenges to graduate schools: The Ph.D. program in ten universities. San Francisco: Jossey-Bass, 1970.
- Hofstadter, R. & Smith, W. (Eds.). American higher education: A documentary history, Vol. 1. Chicago: The University of Chicago Press, 1970.
- Hollis, E. V. Toward improving Ph.D. programs. Washington, D. C.: American Council on Education, 1945.
- Hough, J. B. An observation system for the analysis of classroom instruction. In E. J. Amidon & J. B. Hough (Eds.), Interaction analysis: Hearing, research and application. Reading, Mass.: Addison Wesley, 1967.
- Hough, J. B. An observational system for analysis of classroom instruction. A paper presented at the American Educational Research Association Convention, 1965.
- Hull, N. D. A statistical package for the social sciences (SPSS). New York: McGraw-Hill, 1970.

Individualized inservice teacher education: A performance-based module.
Florida State Department of Education, 1971. ERIC #ED 055-049.

Isaac, S., & Michael, W. B. Handbook in research and evaluation. San Diego: Robert R. Knapp, 1971.

Johnson, G. B. Evaluation of the University of South Alabama College of Education teacher training program by analyses of its alumni, 1969. University of South Alabama, Mobile, 1969. ERIC #ED 062 371.

Johnson, G. B. Evaluation of the University of South Alabama College of Education teacher training program by analyses of its alumni, II - 1971. University of South Alabama, Mobile, 1971. ERIC #ED 060 050.

Kerlinger, F. N. Foundations of behavioral research. 2nd ed. New York: Holt, Rinehart & Winston, 1973.

Mayhew, L. B. Colleges today and tomorrow. San Francisco: Jossey-Bass, 1969.

Mortimer, K. P. Accountability in higher education. Washington, D. C.: American Association for Higher Education, 1972.

National Education Association, Research Division. Appendix to sampling and statistics handbook for surveys in education. Washington, D. C.: National Education Association, 1965.

National Education Association, Research Division. Small sample techniques. NEA Research Bulletin, 1960, 38(4), 99-104.

Napier, L. Implementation of a competency-based teacher education program. Final Report, Vol. I. Mississippi State College, 1971. ERIC #ED 065 657.

Novick, M. R., & Jackson, P. H. Statistical methods for educational and psychological research. New York: McGraw-Hill, 1974.

Quick, R. H. Essays on educational reformers. New York: D. Appleton & Co., 1896.

Roose, K., & Anderson, C. J. A rating of graduate programs. Washington, D. C.: American Council on Education, 1970.

Rudolph, F. The American college and university: A history. New York: Alfred Knopf, 1968.

Sandefur, J. T. An illustrated model for the evaluation of teacher education graduates. Washington, D. C.: American Association of Colleges for Teacher Education, 1970.

- Schwartz, A., & Tiedeman, S. C. Evaluating student progress in the secondary schools. New York: Longmans, Green & Co., 1957.
- Senter, R. J. Analysis of data. Glenview, Ill.: Scott, Foresman & Co., 1969.
- Simon, A., & Bloyer, G. Mirrors for behavior: An anthology of classroom observation instruments. Philadelphia: Research for Better Schools, 1967.
- Smith, E. R., Tyler, R. W., et al. Appraising and recording student progress. Adventure in American Education, Vol. III. New York: Harper & Brothers, 1942.
- Snedecor, G. & Cochran, W. Statistical methods. 6th ed. Ames: Iowa State University Press, 1967.
- Stouffer, F. A. (Ed.). Measurement and prediction. Princeton, N. J.: Princeton University Press, 1950.
- Teachers for tomorrow. Bulletin No. 2. New York: The Fund for the Advancement of Education, 1955.
- Thirty schools tell their story. Adventure in American Education, Vol. V. New York: Harper & Brothers, 1943.
- Tuckman, B. W. Conducting educational research. New York: Harcourt, Brace & Jovanovich, 1972.
- Veldman, D. J. Fortran programming for the behavioral sciences. New York: Holt, Rinehart & Winston, 1967.
- Wilson, L. Foreword. In K. Roose & C. J. Anderson, A rating of graduate programs. Washington, D. C.; American Council on Education, 1970.
- Woodring, P. New directions in teacher education. An Interim Report of the Work of The Fund for the Advancement of Education in the Areas of Teacher Education and Recruitment. New York: The Fund for the Advancement of Education, 1957.
- Woodring, P. A fourth of a nation. New York: McGraw-Hill, 1957.
- Wotring, J. C. A pilot study to evaluate teachers educated at Central Michigan University. Central Michigan University, Mount Pleasant, 1972. ERIC #ED 064 230.

DISSERTATIONS RELATED TO ASSESSMENT PROCEDURE*

Alabama

Beckers, Wilmer Henry. An Evaluation of the Student Personnel Program at the Auburn University School of Education. 1961. Vol. 2210, p. 2298. (Auburn University)

Chambliss, Thomas Anderson. An Evaluation of the Auburn University School of Education Student Teaching Program. 1963. Vol. 2403, p. 1085. (Auburn University)

Ottinger, Richard Estes. An Evaluation of the Auburn University Program of Professional Preparation in Physical Education, 1955-1961. 1963. Vol. 2412, p. 5179. (Auburn University)

Phillips, Ray Cook. An Evaluation of the Cooperative Program for Instructional Improvement. 1961. Vol. 2204, p. 1072. (Auburn University)

Talley, Therman Lavon. An Evaluation of the Program Leading to the Doctor of Education Degree in Educational Administration at Auburn University. 1968. Vol. 2811 A, p. 4432. (Auburn University)

California

Aberle, John Wayne. An Evaluation of a College's Curriculum in Business. 1955. Vol. 1503, p. 378. (Stanford University)

Martinez, Frank Robert. Criteria for the Evaluation of the Curriculum in Adult Education. 1963. Vol. 2403, p. 1062. (University of Southern California)

Colorado

Abrolat, Jr., Gustav Adolph. An Appraisal of a Set of Nationally Developed Standards for the Evaluation of Teacher Education Programs. 1969. Vol. 3010 A, p. 4297. (University of Colorado)

Engel, Arthur. An Evaluation of the Teacher Education Program at Huron College. (Research Study No. 1). 1965. Vol. 2609, p. 5262. (Colorado State College)

*All references are to be found in Dissertation Abstracts.

McKittrick, Max Oller. An Evaluation of Faculties and Facilities for Business Teacher Education in Ohio Colleges and Universities. 1960. Vol. 0022, p. 127. (Colorado State College)

Ruman, Edward L. An Evaluation of a Teacher's College General Education Program in Terms of Changes in Certain Attitudes. 1953. Vol. 0015, p. 184. (Colorado State College)

Connecticut

Magrath, George. An Evaluation of Teachers Trained Under the Intensive Program for College Graduates. 1960. Vol. 2107, p. 1864. (University of Connecticut)

Florida

Hamilton, Marshall. An Evaluation of the Florida State University Internship Program. 1953. Vol. 1306, p. 1100. (Florida State University)

Illinois

Kauss, Theodore Rodney. An Evaluation of the Master of Arts in Teaching Program of Teacher Education at Northwestern University. 1968. Vol. 2907 A, p. 2136. (Northwestern University)

Prince, Joe Norman. An Evaluation of Graduate Music Education Programs at the University of Illinois. 1968. Vol. 2907 A, p. 2295. (University of Illinois)

Indiana

Vineyard, Benny Stinson. An Evaluation of the Industrial Supervision Curriculum at Southern Illinois University. 1962. Vol. 2311, p. 4176. (Indiana University)

Iowa

Van Deventer, Herbert C. An Evaluation of the Effects of the Core Program Upon Those Graduates Majoring in the Social Sciences at the State University of Iowa. (1950-52). 1953. Vol. 1306, p. 10055. (State University of Iowa)

Kansas

Groves, Quentin D. A Guide for the Evaluation of Programs of Physical Education. 1962. Vol. 2402, p. 609. (University of Kansas)

Regier, Herold G. An Evaluation of Teacher Competencies of a Selected Group of Graduates of the School of Education, University of Kansas. 1962. Vol. 2309, p. 3251. (University of Kansas)

Michigan

Lokers, Raymond James. An Evaluation of the Doctoral Program in Education at the University of Michigan in Terms of Fulfillment of the Expectations of the Recipients and of the Expectations of the University. 1959. Vol. 1912, p. 3191. (University of Michigan)

Mississippi

Henry, James Oswald. An Evaluation of a Teacher Education Program by the Graduates of a State College. 1964. Vol. 2604, p. 2067. (University of Mississippi)

Pittman, Kathleen Higdon. An Evaluation of a Teacher Education Program by the Graduates of a State College. 1964. Vol. 2603, p. 1501. (University of Mississippi)

Missouri

Price, Marjorie Ann. An Evaluation of the Undergraduate Professional Preparation Programs in Physical Education in Missouri Colleges and Universities. 1968. Vol. 2908 A, p. 2551. (University of Missouri)

Nebraska

Coon, George E. An Evaluation of the Student Teaching Programs at Wayne State University. 1963. Vol. 2504, p. 2293. (Wayne State University)

Cowing, Donald Earl. Doctoral Graduates in Special Education at Wayne State University--Their Characteristics and Evaluation of the Program. 1969. Vol. 3009 A, p. 3825. (Wayne State University)

Davis, Don E. An Evaluation of a Pre-Service Program of Teacher Education Based Upon the Opinions of In-Service Teachers. 1952. Vol. 1301, p. 36. (Wayne State University)

- Hardt, Walter Frederick. An Evaluation of the Professional Education Program in Two Lutheran Teachers' Colleges. 1961. Vol. 2112, p. 3708. (University of Nebraska)
- Irwin, Claire Cook. The Doctor of Education Program at Wayne State University--An Appraisal of Institutional Aims, Recipient Satisfaction, Faculty Evaluation, and Dissertation Quality. 1960. Vol. 2101, p. 132. (Wayne State University)
- Kirch, Martin Bernard. An Evaluation of the Lutheran Teacher Education Program of Concordia Teacher's College, Seward, Nebraska. 1959. Vol. 2008, p. 3153. (University of Nebraska)
- Lindau, Ora Frank. A Follow-up and Student Evaluation of the Teacher Education Program in Industrial Education at Kearney State College. 1968. Vol. 2906 A, p. 1807. (University of Nebraska)
- Schaefer, Reed Norman. An Evaluation of the Teacher Education Program at Parsons College. 1961. Vol. 2209, p. 3106. (University of Nebraska)
- Simonds, John Wygant. An Evaluation of the Educational Administration Internship Program at the Livonia, Michigan, Public Schools. 1965. Vol. 2609, p. 5182. (Wayne State University)
- Thompson, Gerry E. An Evaluation of the Teacher Education Program of Union College, Lincoln, Nebraska. 1965. Vol. 2601. p. 217. (University of Nebraska)

New York

- Corey, Whitney Thompson. An Evaluation of the Professional Preparation of Men High School Physical Education Teachers at State University College, Cortland, New York. 1965. Vol. 2603, p. 1495. (Syracuse University)
- Harmon, Richard Prescott. An Evaluation of the Teacher Education Program of the School of Education of Long Island University. 1966. Vol. 2707 A, p. 2084. (Fordham University)
- Lastinger, Samuel Thomas. An Evaluation of the Student Teaching Program at Florida Southern College. 1952. Vol. 1301, p. 47. (New York University)
- Lee, Donald Edward. An Evaluation of Certain Objectives of High School Physics with Reference to the Training of Future College Physics Majors and the General Education of All Students . . . 1959. Vol. 0011, p. 3665. (New York University)

Stuart, Alden Taylor. An Evaluation of the Program in Educational Administration at Syracuse University. 1956. Vol. 1608, p. 1390. (Syracuse University)

North Dakota

Grovom, Evelyn Lucile. An Evaluation of the Business Teacher Education Curricula in Colleges Based on the Opinions of the Business Teacher in the Public High Schools in the State of Minnesota. 1968. Vol. 2911 A, p. 3896. (University of North Dakota)

Ohio

Tanner, Daniel. An Evaluation of the Undergraduate Program at the Ohio State University by Agricultural Education Graduates (Volumes I and II). 1955. Vol. 1604, p. 712. (Ohio State University)

Towers, Edward Roy. Industrial Arts Teacher Education--An Evaluation and Projection of the Undergraduate Program of the Ohio State University. 1956. Vol. 1705, p. 1083. (Ohio State University)

Oklahoma

Lindley, Jesse B. The Development of a Teacher Evaluation Program. 1962. Vol. 2305, p. 1580. (University of Oklahoma)

Loman, M. Laverne. An Experimental Evaluation of Two Curriculum Designs for Teaching First-Year Algebra in a Ninth Grade Class. 1961. Vol. 2202, p. 502. (University of Oklahoma)

Vance, Merle William. Evaluation of Teacher Education Programs in the State of Oklahoma. 1955. Vol. 1512, p. 2484. (University of Oklahoma)

Ward, William Wedekind. An Evaluation of the Business Education Program at Southwestern State College. 1956. Vol. 1701, p. 107. (University of Oklahoma)

Wedel, Mack Robert. A Critical Analysis of a Teacher Evaluation Program. 1964. Vol. 2510, p. 5658. (University of Oklahoma)

Oregon

Sathinecker, Robert Eugene. An Evaluation of the Masters Program in Teacher Education at Walla Walla College. 1965. Vol. 2608, p. 4348. (University of Oregon)

Wolfe, Arnold Dale. An Evaluation of the General Studies Degree Program at Southern Oregon College Based on the Opinions of Graduates. 1964. Vol. 2505, p. 2820. (University of Oregon)

Pennsylvania

Stapleton, Martin Luther. An Evaluation of Two Programs of Student Teacher Supervision by College Supervisors. 1965. Vol. 2609, p. 4483. (Pennsylvania State University)

Tennessee

Barkat, Samuel. Development of Criteria and their Employment for Evaluation of Student Teaching Programs Offered by the Colleges and Universities in Tennessee. 1963. Vol. 2411, p. 4560. (University of Tennessee)

Lucke, Edward James. An Evaluation of the Professional Preparation Program in Health and Physical Education at Lenoir Rhyne College. 1963. Vol. 2412, p. 5177. (George Peabody College for Teachers)

Texas

Kromminga, Louis Ude. Characteristics and Curriculum Evaluations of Bachelor and Master of Music Degree Graduates from the University of Texas (1948-1957). 1959. Vol. 2002, p. 575. (University of Texas)

Utah

Hinckley, Ira Lee. An Evaluation of the Teacher Education Program at Illinois Teacher's College Chicago-North. 1967. Vol. 2804 A, p. 1327. (University of Utah)

Virginia

Trausneck, William McLean. Development of an Instrument for the Self-Evaluation of School Health Service Programs. 1963. Vol. 2412, p. 5156. (University of Virginia)

Washington

Chatburn, Acel Handy. An Evaluation of the Program of Boise Junior College by its Graduates. 1956. Vol. 1701, p. 68. (Washington State University)

Goodwin, Luther. An Evaluation of Teacher Education in the Physical Education Degree Program at the University of Alberta. 1962. Vol. 2308, p. 2781. (University of Washington)

DIRECTORY OF INSTITUTIONS
ASSOCIATED WITH
TEACHER PREPARATION ASSESSMENT PROGRAMS

Alabama

University of South Alabama; Granville B. Johnson, 1969 and 1971.

Arizona

Arizona State University; Dan Kauffman, 1973.
Northern Arizona University; Sam F. McClanahan, 1973.
University of Arizona; Roy M. Claridge, 1969.

Colorado

Colorado State College; Max Oller McKittrick, 1960.

Illinois

Illinois Teacher's College, Chicago-North; Ira L. Hinckley, 1967.

Indiana

Indiana University, Jerome C. Harste, 1972.

Iowa

Parsons College; Reed Schaefer, 1962.

Louisiana

McNeese State University; Kathleen H. Pittman, 1964.

Michigan

Central Michigan University; Jarvis C. Wotring, 1972.

Minnesota

University of Minnesota; Robert T. Alciatore and Ruth E. Eckert,
1968.
University of Minnesota; Jerome C. Harste, 1972.

Mississippi

Mississippi State College; Lee Napier, 1971.

New York

Columbia University; Bruce R. Joyce, et al., 1971.

South Dakota

Huron College; Arthur Engel, 1965.

Virginia

Old Dominion University; James O. Henry, 1964.